FLAW DETECTION IN ALUMINIUM CASTINGS LEVERAGING SYNTHETIC DATA FOR NON-DESTRUCTIVE TESTING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

UMUT CAN ERKAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MODELING AND SIMULATION

SEPTEMBER 2025

FLAW DETECTION IN ALUMINIUM CASTINGS LEVERAGING SYNTHETIC DATA FOR NON-DESTRUCTIVE TESTING

submitted by **UMUT CAN ERKAN** in partial fulfillment of the requirements for the degree of **Master of Science in Modeling and Simulation Department, Middle East Technical University** by,

Date: 01.09.2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Umut Can Erkan

Signature :

ABSTRACT

FLAW DETECTION IN ALUMINIUM CASTINGS LEVERAGING SYNTHETIC DATA FOR NON-DESTRUCTIVE TESTING

Erkan, Umut Can
M.S., Department of Modeling and Simulation
Supervisor: Assoc. Prof. Dr. Elif Sürer
Co-Supervisor: Prof. Dr. Ulaş Yaman

September 2025, 70 pages

Non-destructive testing (NDT) using radiographic imaging plays a vital role in quality assurance for industrial manufacturing, particularly in detecting internal flaws in aluminum castings. However, the application of deep learning to radiographic flaw detection is limited by a lack of publicly available, diverse datasets and domain-specific pretraining strategies. This thesis addresses both challenges by introducing a high-resolution X-ray dataset of aluminum castings annotated according to the American Society for Testing and Materials (ASTM) standards. The dataset covers both coarse and fine-grained flaw categories, such as porosity types, gas holes, inclusions, and shrinkages, and captures variation across materials, perspectives, and flaw severity levels. We benchmark several object detection models on this dataset to establish baseline performance and highlight domain-specific challenges. To overcome data scarcity, we propose a synthetic data generation pipeline using Stable Diffusion with textual inversion. Out-of-distribution synthetic radiographs are generated from only a few publicly available reference images, ensuring reproducibility and eliminating dataset-specific leakage. The resulting synthetic data is used for self-supervised contrastive pretraining, enabling the model to learn defect-aware representations without requiring labeled data. Our pretrained backbone improves the performance of downstream tasks on our dataset, outperforming ImageNet-pretrained baseline. This thesis' main contributions are a challenging new dataset, a dataset-agnostic synthetic data generation pipeline, and empirical evidence that synthetic, out-of-distribution samples can still drive effective self-supervised learning for industrial flaw detection.

Keywords: casting defect detection, non-destructive testing, self-supervised learning, synthetic data

ÖZ

TAHRİBATSIZ MUAYENE İÇİN SENTETİK VERİLER KULLANILARAK ALÜMİNYUM DÖKÜM KUSURLARININ TESPİT EDİLMESİ

Erkan, Umut Can
Yüksek Lisans, Modelleme ve Simülasyon Bölümü
Tez Yöneticisi: Doç. Dr. Elif Sürer
Ortak Tez Yöneticisi: Prof. Dr. Ulaş Yaman

Eylül 2025, 70 sayfa

Endüstriyel üretimde kalite güvencesi açısından radyografik görüntüleme ile yapılan tahribatsız muayene (NDT), özellikle alüminyum dökümlerindeki iç kusurların tespiti açısından kritik bir rol oynamaktadır. Ancak, derin öğrenmenin radyografik kusur tespitine uygulanması, alana özgü ön eğitim stratejilerinin ve halka açık, çeşitli etiketli veri kümelerinin eksikliği nedeniyle sınırlı kalmaktadır. Bu tez, her iki sorunu da ele alarak, American Society for Testing and Materials (ASTM) standartlarına göre etiketlenmiş yüksek çözünürlüklü alüminyum döküm X-ışını görüntülerinden oluşan yeni bir veri kümesi sunmaktadır. Veri kümesi, porozite türleri, gaz boşlukları, inklüzyonlar ve çekintiler gibi hem kaba hem de ince taneli kusur kategorilerini kapsamaktadır. Aynı zamanda malzeme türü, görüntüleme perspektifi ve kusur şiddeti açısından çeşitlilik içermektedir. Bu veri kümesi üzerinde çeşitli nesne tespiti modelleri değerlendirilerek, temel performans ölçütleri belirlenmiş ve alana özgü zorluklar ortaya konmuştur. Etiketli veriye duyulan gereksinimi azaltmak için, Textual Inversion ile yönlendirilen Stable Diffusion tabanlı sentetik veri üretim hattı önerilmiştir. Sadece birkaç halka açık referans görüntüden yola çıkılarak, dağılım dışı sentetik radyografiler üretilmiştir. Bu sayede yöntem tekrar üretilebilir hale gelmiş ve veri kümesine özgü sızıntılar engellenmiştir. Üretilen sentetik görüntüler, kendinden gözetimli karşıt öğrenme yöntemiyle modelin ön eğitiminde kullanılmıştır. Elde edilen ön eğitimli omurga modeli, hem top-1 doğruluğunda hem de kusur tespiti görevinde, ImageNet ile ön eğitilmiş karşılaştırmalı modeli geride bırakmıştır. Bu tez, döküm NDT alanı için zorlu bir yeni veri kümesi, veri kümesinden bağımsız bir sentetik veri üretim hattı ve dağılım dışı sentetik verilerle etkili temsillerin öğrenilebileceğine dair deneysel kanıtlar sunmaktadır.

Anahtar Kelimeler: döküm kusur tespiti, tahribatsız muayene, kendinden gözetimli öğrenme, sentetik veri

To my family

ACKNOWLEDGMENTS

I express my gratitude to my advisor Assoc. Prof. Dr. Elif Sürer and co-advisor Prof. Dr. Ulaş Yaman for their supervision.

I also thank Dr. Tuğçe Kaleli Alay, members of METU KATAMER, and ASELSAN for their support throughout this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ASTM	American Society for Testing and Materials
CLIP	Contrastive Language–Image Pretraining
COCO	Common Objects in Context
CNN	Convolutional Neural Network
FPN	Feature Pyramid Network
DINO	Self-Distillation with No Labels
DR	Digital Radiography
FAISS	Facebook AI Similarity Search
LAION	Large-scale Artificial Intelligence Open Network
IoU	Intersection over Union
mAP	Mean Average Precision
NDT	Non-Destructive Testing
OOD	Out-of-Distribution
PQ	Product Quantization
SSL	Self-Supervised Learning
ViT	Vision Transformer
T2I	Text-to-Image
R2S	Real-to-Synthetic

CHAPTER 1

INTRODUCTION

In the manufacturing industry, non-destructive testing (NDT) plays a vital role in ensuring the structural integrity and reliability of components without causing damage. These techniques are particularly critical in high-stakes sectors such as aerospace, automotive, and heavy machinery, where undetected flaws can lead to catastrophic failures. Among the various NDT methods, radiographic inspection using X-ray imaging is one of the most widely adopted approaches for identifying internal flaws in castings, including porosity, cracks, shrinkages, gas holes, and inclusions [1]. Aluminum castings, in particular, are extensively used in industrial applications due to their combination of low weight, corrosion resistance, and mechanical strength [2].

Despite its widespread use, radiographic analysis in industrial settings is still predominantly carried out by trained human inspectors. This process is inherently manual, subjective, and labor-intensive. Moreover, it is prone to inconsistencies between inspectors and can be affected by fatigue, leading to variability in flaw interpretation and missed defects [3]. These limitations have caused growing interest in automating flaw detection through computer vision and deep learning technologies.

While deep learning has revolutionized many computer vision applications, its adoption in radiographic flaw detection has been slow due to several domain-specific challenges. The most important among them is the lack of publicly available, large-scale, standardized and annotated datasets that accurately represent real-world industrial inspection conditions. Most available datasets either contain data focusing on few materials and perspectives or lack the fine-grained flaw annotations necessary for reliable evaluation.

Moreover, unlike natural images, radiographs present domain-specific complexities such as subtle grayscale variations, varying material thicknesses, and low-contrast flaw structures. This makes feature learning more difficult and limits the effectiveness of transfer learning from conventional datasets such as ImageNet [4].

To address these challenges, this thesis investigates automated casting flaw detection using high-resolution X-ray images by introducing a novel dataset and proposing a synthetic data-driven self-supervised pretraining framework. The core idea is to reduce reliance on large annotated datasets by leveraging synthetic radiographs created using only a few publicly available reference images. Through domain-aware synthetic generation and a contrastive learning objective, we aim to develop a robust backbone model that increases the performance on real NDT data, even in the presence of limited annotations.

1.1 Research Questions

This study is guided by the following key research questions:

1. Can we build a standardized and diverse benchmark dataset for flaw detection that reflects industrial casting scenarios and enables comprehensive evaluation of detection models?

2. How do different object detection models perform on the proposed benchmark, and what insights can be drawn regarding their generalizability and limitations in industrial inspection settings?

3. How can synthetic radiographic data be generated to simulate realistic and class-specific casting flaws without leaking private data or requiring large amounts of labeled examples?

4. Can self-supervised learning benefit from these synthetic samples, particularly when the flaws represent out-of-distribution concepts not present in large-scale diffusion models?

5. How does pretraining on synthetic flaw radiographs affect the performance of downstream tasks such as flaw classification and detection on the proposed dataset?

1.2 Motivation

The main motivation for this study stems from the need for reliable AI tools in industrial inspection pipelines. Manual flaw detection is not only costly and time-consuming, but also prone to human error, especially when dealing with subtle or fine-grained defects. While deep learning holds promise, most existing approaches suffer from a limitation: lack of domain-specific data.

Radiographic images differ significantly from natural images in terms of their low contrast, illumination invariant and grayscale-only appearance. Flaws are often subtle, low signal-to-noise, and sensitive to variations in exposure, positioning, or material thickness. These characteristics reduce the transferability of features learned from general image datasets like ImageNet [4], and emphasize the need for domain-specific training data.

At the same time, modern generative models such as Stable Diffusion [5] offer an opportunity to simulate images with only few references, thereby eliminating the need to collect or label large-scale real data. However, such synthetic data must be realistic, diverse, and class-distinctive to be effective in pretraining. By combining these capabilities with self-supervised contrastive learning, we aim to create a robust flaw detection pipeline that can learn meaningful visual representations purely from synthetic data, and increase the performance on our dataset by bridging the domain gap.

1.3 Contributions of the Study

This thesis makes several key contributions to the field of automated flaw detection and industrial NDT:

- **A New Flaw Detection Dataset for Castings:** We introduce a radiographic dataset containing high-resolution, high bit-depth images of aluminum castings with annotations following ASTM standards [6]. The dataset includes both coarse and fine-grained flaw categories, variety of materials and perspectives, and covers a broad spectrum of flaw grades which are critical for industrial use. We evaluate several object detection models on the proposed dataset to investigate their ability to generalize to the specific challenges of radiographic flaw detection in castings.
- **Synthetic Flaw Generation with Few-Shot Embeddings:** We propose to use a framework that uses synthetic radiographs generated via Stable Diffusion, guided by textual inversion embeddings learned from only a few publicly available reference images. This approach ensures full reproducibility and prevents dataset-specific leakage.
- **Self-Supervised Representation Learning with Synthetic Radiographs:** We propose a representation learning framework specifically designed for radiographic inspection of industrial defects. We analyze whether self-supervised learning (SSL) benefits from synthetic images generated from out-of-distribution concepts that are introduced solely through textual inversion with a few reference images. Specifically, we investigate whether using data not available in the original Stable Diffusion training set, such as casting defects, can still lead to effective visual representations in SSL. We further evaluate how this knowledge transfers to downstream tasks on our dataset for non-destructive testing. Unlike the only prior SSL study in this domain [7], which uses global instance-level similarity on real unlabeled images, our approach leverages object-level flaw crops and class conditioning, enabling finer-grained feature learning.

1.4 Organization of the Thesis

This thesis is organized into six chapters, each addressing a distinct component of the research conducted in the context of automated casting flaw detection using radiographic imaging.

- **Chapter 1 – Introduction:** Introduces the research context, formulates the problem, outlines the motivation, research questions, and summarizes the key contributions of the study.
- **Chapter 2 – Related Work:** Reviews existing literature on non-destructive testing, radiographic flaw detection, synthetic data generation, and self-supervised representation learning.
- **Chapter 3 – Methodology:** Describes the technical approach in detail, including the dataset construction, detection benchmarks, synthetic image generation strategies, and the self-supervised learning framework used in this study.

- **Chapter 4 – Implementation and Experiments:** Presents implementation details, dataset statistics, experimental setup, and the results of evaluations.
- **Chapter 5 – Discussion:** Interprets the empirical findings, and highlights the strengths, limitations, and practical relevance of the proposed methods in the NDT domain.
- **Chapter 6 – Conclusion and Future Work:** Summarizes the main findings, states the impact of the contributions, and outlines promising directions for future research.

CHAPTER 2

RELATED WORK

This chapter reviews how casting defect detection in radiographic inspection evolved over the years, followed by examining the role of synthetic data for both casting and welding. It further explores transfer learning strategies in the field and it steps beyond the domain to survey broader computer vision research on integrating synthetic images to self-supervised learning.

2.1 Radiographic Casting Defect Detection

Over the years, there have been many attempts to automatically detect defects in casting materials using radiographic imaging. Earlier work focuses on traditional image processing techniques, statistical modeling, and hand-crafted features combined with machine learning classifiers. Some of the methods include fuzzy logic–based techniques [8], wavelet transform-based methods [9], and thresholding techniques with morphological operations [10] [11]. Statistical modeling such as [12], and a grayscale arranging pairs (GAP) based method [12] have also been explored.

To investigate manual feature extractors, extensive comparative analyses were performed using algorithms including Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Binary Robust Invariant Scalable Keypoint (BRISK) and others while Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN) and Support Vector Machines (SVM) variants were utilized for classification [13]. Moreover, Histogram of Gradients (HOG), Local Binary Features (LBF), and Gabor filtering were also compared in another study [14] combined with classifiers such as SVM and Stochastic Gradient Descent (SGD).

With the breakthroughs in computer vision and object detection, many recent studies have adopted deep learning methods for casting defect detection due to their improved performance over traditional techniques. Du et al. [15] adopted the Feature Pyramid Network (FPN) and RoIAlign to increase the accuracy on their dataset, which comprised binary-labeled defects without specific type annotations. In their follow-up study [16], they integrated DetNet [17] and the Path Aggregation Network (PANet) [18] to address the issue of loss of location information and feature imbalance on feature maps. To further enhance detection performance, they applied soft Non-Maximum Suppression (soft-NMS) to retain high-confidence proposals.

Duan et al. [19] used guided filtering as a preprocessing step while improving YOLOv3 [20] architecture by integrating dual-density convolutional layers. Their dataset consists of two types of defects (blowholes and porosity) with five different levels of looseness. Jiang et al. [21] introduced weakly-supervised learning approach utilizing a VGG-16-based model with additional attention mechanisms to generate attention maps to guide data augmentation. Their work also incorporates mutual-channel loss function to enhance true defect features. Xue et al. [22] replaced YOLOv3's original Darknet-53 backbone with an EfficientNet [23] variant, which both boosts average precision and shrinks model size. The study integrated depth-separable convolutions to further reduce the model parameters. The results were reported on binary-labeled defects without type annotations.

Garcia-Pérez et al. [24] worked on the GDXray dataset [25] of aluminium castings and performed the ablation study of hyper-parameters utilizing a RetinaNet [26] based model. Cheng et al. [27] extended Cascade R-CNN [28] by integrating deformable convolutions into each stage of the cascade, and a spatial-attention module before each region-proposal network (RPN). Their study outperformed standard Cascade R-CNN and Faster R-CNN [29] baselines in locating six types of defects in wheel hubs. Parlak and Emel [30] benchmarked several YOLOv5 [31] variants (n, s, m, l) to balance speed and accuracy, finding that YOLOv5m delivered the best trade-off. They released Al-Cast dataset including shrinkage and gas hole defects.

Wang et al. [32] proposed different extensions to the YOLOv8 [33] model. They preprocessed every image with an adaptive contrast enhancement step that is based on standard deviation to boost low-contrast defects. A frequency domain attention module is proposed to learn the weighing of both low-frequency and high-frequency edges. The features fused at multiple scales through a progressive feature refinement module. They also introduce a modified deep-regression loss (MDPIoU) for better bounding-box alignment and boost the performance on GDXRay dataset. Hai et al. [34] used Gain-Adaptive Multi-Scale Retinex as the preprocessing algorithm and combined FPN with Convolutional Block Attention Module (CBAM). Their work also integrated Weighted Region of Interest (W-RoI) Pooling instead of RoIAlign to avoid area misalignment of detections. Their dataset focuses primarily on blowhole defects.

Inspired by YoloV8 [33] model and the Bidirectional Feature Pyramid Network (BIFPN) [23], Zhang et al. [35] proposed a network that uses a dual-layer encoder–decoder with Hadamard-product fusion to preserve fine defect features. They also replaced the standard CIoU loss with a Normalized Wasserstein Distance (NWD) loss to improve small-box localization on their binary labeled defect dataset.

While CNNs have dominated casting defect detection in X-ray images, the success of transformers in achieving state-of-the-art results in object detection has sparked interest in this direction. However, only a couple of studies have so far applied transformer-based models. For example, Cheng et al. [36] presented AGD-DET, a streamlined DETR-based [37] detector tailored for turbine-blade castings in X-ray radiographs. They utilized a lightweight CNN stem before feeding to a transformer equipped with a Global Dynamic Attention module that adaptively samples both small and large defect regions. A scale-adaptive tokenizer then adjusts each token's receptive field resulting in a 7% boost in mean average precision (mAP) compared to the DETR baseline in their binary-labeled dataset. Similarly, Zhang et

al. [38] adapted the DETR framework for turbine-blade X-ray inspection. This study applied channel-wise attention to highlight informative features and employed a focal-style loss to balance easy and complex examples. The dataset consists of six different defect types and collected using conventional film radiography.

Although X-ray imaging is essential for detecting internal defects in castings, several studies focus solely on surface defects, leveraging other modalities such as conventional imaging [39], [40], [41], [42] or photo-stereoscopic imaging [43]. These approaches, however, are limited to surface-level flaws, as they cannot capture internal flaws like those revealed by radiographic imaging.

There are limited studies classifying defects and mentioning their grades based on the American Society for Testing and Materials (ASTM) standards. Hena et al. [44] focused solely on developing an automated flaw-grading pipeline that uses ASTM 2973-15 to assign four severity levels via k-d trees after detecting flaws in digital X-rays. This study only works on a segmented flaw masks, assumes a fixed 700 mm² area on the images, and tests on the synthetic images which severely limits its capacity of general use. On the other hand, Parlak and Emel [45] proposed a two-stage pipeline, similar to their previous work [30], utilizes YOLOv5 detector and segments the detected defects using clustering based thresholding. This work utilize ASTM E2422 [6] standard for defect classification and grading. However, there is no evaluation for grading as the dataset lacks ground truth data. The proposed Al-Cast dataset extension (Al-Cast2), which introduces an additional porosity class, has not been made publicly available to date.

2.2 The Use of Synthetic Data in Casting Defect Detection

Data scarcity and class imbalance are persistent challenges in casting defect detection, since collecting and annotating large volumes of radiographs with diverse defect types is time-consuming and costly. To address these issues, many researchers have turned to simulation-based generation and augmentation. Early work by Mery et al. [46] projected CAD-modeled cracks and blow-holes onto defect-free radioscopic images using the exponential X-ray attenuation law, and more recent studies have followed that lead. For instance, Mery et al. [47] evaluated YOLO, RetinaNet, and EfficientDet variants after injecting simulated ellipsoidal defects to defect-free X-ray images. The simulation randomly samples ellipse parameters of size, orientation, and position to mimic typical defects. These shapes are inserted into the image by reducing pixel intensity within the ellipse to simulate X-ray attenuation. More recently, Andriiashen et al. [48] used a Monte Carlo engine to generate synthetic X-rays both with and without scatter, showing that including realistic scatter noise is crucial for detecting the smallest defects. Bosse et al. [49] introduced a physics-driven simulation to generate highly realistic porosity defects. They started by embedding CAD-modeled pore geometries of varying sizes and shapes into virtual aluminimum casting volumes, then used a ray-tracing engine to produce synthetic X-ray projections with exact pixel-perfect masks. The trained semantic-segmentation model solely on these synthetic porosity images were evaluated on real industrial radiographs and showed satisfying results.

Learning-based methods, particularly those utilizing generative models, have emerged as a powerful tool for synthetic data generation in casting defect detection. These approaches generate realistic defect images by learning the underlying patterns from existing radiographic datasets. Recently, the advances of generative models replaced hand-crafted simulations with fully learned defect generation. In their study, Goswami et al. [50] presented a two-stage Generative Adversarial Network (GAN) [51] pipeline for generating automotive wheel spokes with porosity defects. In the first stage, the network learns to convert given edge maps into defect-free radiographs. In the second stage, it inpaints pores into those images to user-defined locations. The synthetic images achieve a Structural Similarity Index (SSIM) [52] above 0.90 and closely match real data in gray-level histograms. For evaluation, they run all outputs through an off-the-shelf ISAR defect-detection tool, which successfully identifies 54 percent of synthetic images at an intersection-over-union (IoU) threshold of 0.25. Garcia-Pérez et al. [53] introduced Scalable Conditional Wasserstein GAN (SCWGAN) by integrating a lightweight attention block and a multi-scale feature loss. Their data generation pipeline takes a flawless baseline image with a foreground mask, which is generated through thresholding. The trained GAN model generates three types of defects (porosity, shrinkage, and inclusion). These defects are then inpainted into selected locations and post-processed by adjusting the brightness level and adding Gaussian or Poisson noise. This method boosts the mAP by around 12 percent on GDXRay dataset compared to training on real data alone.

2.3 The Use of Synthetic Data in Welding Defect Detection

Radiographic welding defect detection faces the same problems as casting defect detection. Therefore, it is possible to see attempts to overcome issues of data scarcity and class imbalance utilizing a synthetic data generation approach.

Guo et al. [54] tackled the severe class imbalance in weld-X-ray datasets by using a Contrast Enhancement Conditional GAN (CEcGAN) as a global resampling tool. First, they boost low-contrast images via histogram equalization inside the GAN's generator, ensuring defect and background features both stand out. The conditional branch then learns to produce radiographs with defects such as crack, lack-of-fusion, lack-of-penetration, porosity, and slag inclusion, by conditioning on defect class labels. The results demonstrated that the method outperforms traditional augmentations in defect recognition.

Li et al. [55] employed a simple copy-paste augmentation technique using real defect instances and defect-free samples on their custom dataset. They showed that training with their synthetic samples increases object-detection performance beyond classical augmentations.

Wang et al. [56] utilized a two-channel attention-guided DCGAN [57] both round and strip-type flaws by conditioning on defect classes. The generated samples are used to train the proposed segmentation network which incorporates multi-scale convolutions (combining 2×2 and 4×4 kernels) and embeds a CBAM module within the encoder layers. To validate GAN's output quality, metrics such as Fréchet Inception Distance (FID) [58], Learned Perceptual Image Patch Similarity (LPIPS) [59] and Multiscale Structural Similarity Index Measure (MS-SSIM) [60] were weighed to align with expert ratings.

Recently, Zhang et al. [61] also proposed a DCGAN variant to produce four distinct weld-defect types and improved FID metric compared to the base model. The synthetic defects combined with real defects showed improved classification accuracy on their lightweight MobileNet [62] variant (DG-MobileNet) which incorporates dilated convolutions, DropBlock regularization [63], and Squeeze and Excitation (SE) attention [64].

2.4 Transfer Learning in Radiographic NDT

In areas such as industrial radiographic inspection, obtaining large labeled datasets is often difficult. A common solution to the heavy data demands is using transfer learning, where a model pretrained on a large dataset such as ImageNet [4] is fine-tuned on a smaller, domain-specific dataset. In many downstream tasks, these pretrained backbones converge faster and often outperform models trained from scratch. Similarly, for the radiographic NDT, Ferguson et al. [65] showed ImageNet pretrained model outperforms the scratch model, and COCO [66] pretrained model outperforms both for detection and segmentation. However, the large domain gap between natural images and industrial radiographs means that many of the learned features transfer poorly, as the models remain biased toward edges, colors, and textures common in natural images, rather than the subtle density gradients and noise patterns characteristic of X-rays. Furthermore, natural images vary wildly in semantic content, whereas radiographs look mostly uniform.

Some recent work has tried to close the gap by pretraining on datasets that more closely resemble industrial radiographs. Yu et al. [67] showed that a network first pretrained on medical X-ray images yielded even better transfer performance, since it already knew to recognize local density gradients and noise patterns. Despite these gains, medical X-rays still differ from industrial radiographs in geometry, materials, and imaging protocols, so the improvements remain limited. And to date, there is still no large, variant, and publicly available dataset of defected material radiographs, which continues to bottleneck progress in this field.

Self-supervised learning (SSL) is beneficial in this context because it can learn rich visual features from unlabeled images and reduce reliance on costly human annotations. It is possible to see an example study of Intxausti et al. [7] where researchers use SSL methods such as SimSiam [68] and SimMim [69] on their unlabeled dataset for pretraining. This backbone outperformed ImageNet in terms of detection performance on both their private dataset and GDXRay using Faster-RCNN. However, even collecting unlabeled radiographs can be challenging because of the industrial privacy concerns.

2.5 Beyond Radiographic NDT: Self-supervised Learning with Synthetic Images

In recent years, SSL has revolutionized computer vision by learning feature representations that rival or even outperform those from fully supervised models. Techniques such as contrastive learning (SimCLR [70], MoCo [71]) and masked image modeling (MAE [72], SimMIM [69]) teach networks to pull together different augmented views or to reconstruct miss-

ing patches, all without labels. However, these methods still depend on vast collections of unlabeled images.

Although the use of synthetic images are mostly explored in supervised settings [73] [74] [75], there are recent works that combined synthetic images with SSL methods. That is because recent advancements in generative models, particularly diffusion models, have enabled the creation of highly realistic synthetic images, which can be used as a resource for pretraining models. For instance, Tian et al. [76] pioneered a self-supervised framework (StableRep) that learns representations from synthetic images that were generated by a text-to-image diffusion model Stable Diffusion [5]. The study used image caption datasets and used the captions as text prompts to generate multiple images from the same text prompt. Then, these generated images are treated as positive samples, rather than treating just one augmented view as positive. In other words, given K images sampled per prompt, each image uses the other K–1 as positives in the contrastive loss which increases the number of positive pairs and enriching learned invariances. This work matches or exceeds SSL baselines trained on real images such as SimCLR and MoCo and even outperforms CLIP [77] trained on larger real datasets. They also report better downstream results on semantic segmentation and few-shot detection than existing methods. The follow-up work SynCLR [78] uses Large Language Models (LLMs) to generate the captions instead of using real image caption datasets. The rest of the implementation details remains the same. The results improved the previous work even further.

More recently, Bafghi et al. [79] presented MixDiff that combines real and synthetic images in pretraining stage. Instead of pretraining on only synthetic or only real images, MixDiff pairs augmented view of a real image with a synthetic version using the image-to-image generation via Versatile Diffusion [80], an off-the-shelf variant of Stable Diffusion. The aim is to take advantage of the diversity of synthetic data but also ground it in real image statistics to bridge the gap that purely synthetic methods have. MixDiff can be combined with joint-embedding SSL methods such as SimCLR [70], BarlowTwins [81], and DINO [82]. When plugged into SimCLR on ImageNet-1K, it raises top-1 accuracy by 4.56 percent over the standard two-view setup.

Considering the latest advancements, leveraging synthetic data in the context of SSL stands out as a promising area of research for industrial radiographic inspection.

CHAPTER 3

METHODOLOGY

3.1 Stable Diffusion

Essentially, a diffusion model gradually adds noise to real images over T timesteps, then learns to reverse that process. Stable Diffusion [5], however, operates in a latent space learned by a variational autoencoder (VAE) [83]. This reduces computational requirements significantly. Therefore, the latent diffusion framework enhances efficiency for high-resolution image generation.

The pretrained encoder $\mathcal{E}$ maps an image x to a latent $z = \mathcal{E}(x)$, where $z \in \mathbb{R}^{h \times w \times d}$. Typically $h = H/8, w = W/8, d = 4$. The decoder $\mathcal{D}$ reconstructs the image: $\hat{x} = \mathcal{D}(z)$. The VAE was trained to minimize:

$$\mathcal{L}_{\text{auto}} = \|x - \mathcal{D}(\mathcal{E}(x))\|_2^2 + \lambda \text{KL}(q(z|x)\|p(z)). \tag{1}$$

Here, the term $q(z|x)$ is the encoder's distribution over latents given the image, while $p(z)$ is a standard Gaussian prior. The KL divergence ensures $q(z|x)$ resembles $p(z)$, regularizing the latent space.

Once in the latent space, a diffusion process is applied. The forward process adds noise over T timesteps:

$$q(z_t \mid z_{t-1}) = \mathcal{N}\big(z_t;\, \sqrt{1-\beta_t}\, z_{t-1},\, \beta_t I\big) \tag{2}$$

where β_t is a fixed variance schedule and I is the identity matrix. The reverse process, learned by a denoising U-Net [84], approximates:

$$p_\theta(z_{t-1} \mid z_t) = \mathcal{N}\big(z_{t-1};\, \mu_\theta(z_t, t),\, \beta_t I\big) \tag{3}$$

The U-Net predicts noise minimizing a mean squared error loss which corresponds to optimizing μ_θ to denoise z_t.

To guide image generation, Stable Diffusion incorporates text prompts using a pretrained CLIP text encoder [77]. Given a prompt, the encoder produces a set of embeddings $\tau = \{\tau_i\}_{i=1}^{M}$, where each $\tau_i \in \mathbb{R}^D$. These embeddings are integrated into the U-Net through

cross-attention layers. Specifically, the attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{4}$$

where the queries Q are derived from intermediate U-Net features, and keys K and values V are computed from the text embeddings. This enables the model to align generated images with the semantic meaning of the prompt.

To improve the alignment between the generated images and the conditioning text, Stable Diffusion employs classifier-free guidance [85]. This technique enhances prompt-guided denoising by blending conditioned and unconditioned noise predictions.

During training, the text conditioning is randomly dropped with probability p_{drop} (typically 0.1). Therefore, the model learns to generate images both with and without text guidance. For conditional generation with prompt c, the model predicts $\epsilon_\theta(z_t, t, c)$. For unconditional generation, it predicts $\epsilon_\theta(z_t, t, \emptyset)$, where $\emptyset$ represents the null text condition. At inference time, classifier-free guidance combines these predictions using a guidance scale w:

$$\tilde{\epsilon}_\theta(z_t, t, c) = \epsilon_\theta(z_t, t, \emptyset) + w \cdot (\epsilon_\theta(z_t, t, c) - \epsilon_\theta(z_t, t, \emptyset)) \tag{5}$$

Here, $w > 1$ amplifies the difference between conditional and unconditional predictions. Higher values of w produce images that more closely follow the text prompt while sacrificing diversity. Thus, w provides a trade-off between diversity and quality.

Overall, the architecture of Stable Diffusion combines a pretrained autoencoder, a denoising U-Net, a text-conditioned cross-attention mechanism, and classifier-free guidance for precise control. Fig. 1 illustrates the overall architecture.

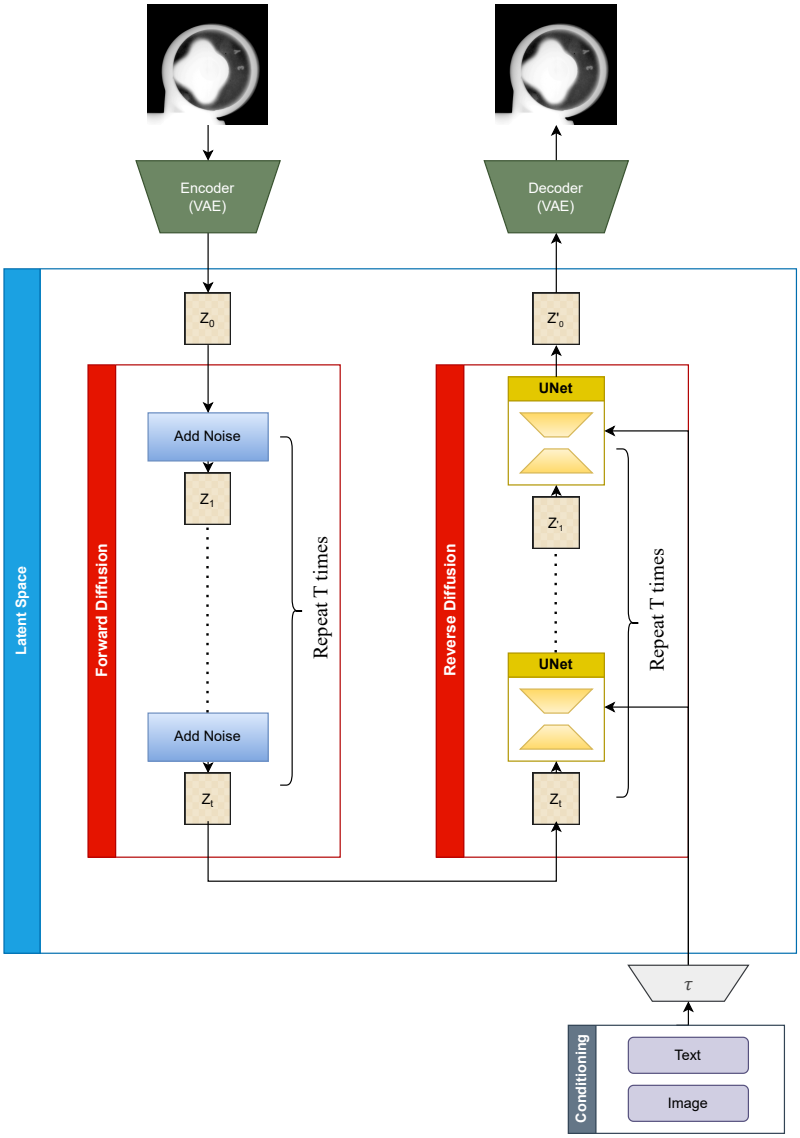


Figure 1: Architecture of Stable Diffusion

3.2 Textual Inversion

Textual Inversion [86] is a technique for introducing new concepts to text-to-image models using only a few example images. Instead of fine-tuning the entire model which requires extensive data and computation, it learn a new token embedding that represents a specific concept, such as a unique object or style. This concept is associated with a pseudo-word S^*, which can then be used in prompts to guide image generation.

During training, given a few samples $\{x_i\}_{i=1}^{N}$ representing a concept, textual inversion optimizes only the embedding vector $v^* \in \mathbb{R}^D$ corresponding to S^*. The process uses the same diffusion loss as standard training. However, it backpropagates gradients only to v^*, keeping all other parameters frozen. Thus, the optimization problem becomes finding v^* that minimizes the expected denoising loss when S^* appears in prompts.

$$\mathcal{L}_{\text{textual}} = \mathbb{E}_{t,\epsilon\sim\mathcal{N}(0,I)}\left[\|\epsilon - \epsilon_\theta(z_t, t, \tau(v^*))\|_2^2\right] \tag{6}$$

Here, z_t is the noisy latent, ϵ is sampled noise, and ϵ_θ is the U-Net prediction conditioned on $\tau(v^*)$, the embedding.

3.3 Multi Positive Contrastive Learning

One dominant self-supervised learning (SSL) approach is contrastive learning, which trains an image encoder to distinguish between similar (positive) and dissimilar (negative) image pairs. Usually, positive pairs are created using traditional augmentations of the same image, while negatives are drawn from other images in the batch. Popular frameworks such as SimCLR [70] and MoCo [71] follow this setup. However, this two-view model limits the variety of positives. To overcome this, multi positive contrastive learning was proposed [76]. In this setting, a text-to-image model generates multiple images from the same caption using different random seeds. The generated images from the same prompt can be treated as positives with respect to one another.

Given a batch of captions, each yielding K synthetic images, an image encoder f_θ extracts embeddings $\{z_{j,1}, \ldots, z_{j,K}\}$ for each group j. The contrastive loss is computed so that each embedding is pulled closer to the other $K-1$ embeddings from the same group, and pushed away from embeddings belonging to other prompts.

Formally, the multi-positive contrastive loss is defined as:

$$\mathcal{L}_i = -\frac{1}{|P(i)|}\sum_{p\in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a\neq i}\exp(z_i \cdot z_a/\tau)} \tag{7}$$

where $P(i)$ is the set of other images generated from the same caption (the positive set for anchor z_i), and τ is a temperature scaling factor. This formulation allows capturing richer intra-prompt variation across multiple positives.

3.4 Proposed Method

The aim of this method is to investigate the effectiveness of self-supervised learning for industrial flaw detection using diffusion models. Specifically, we focus on radiographic casting inspection and explore whether few-shot image synthesis using only publicly available ASTM reference images can support self-supervised representation learning. For pretraining, we integrate multi-positive contrastive loss and synthetic images to enhance learned representations. Ultimately, the goal is to investigate the transferability of these representations to the

downstream casting NDT tasks. Moreover, by collecting a comprehensive dataset of flaws, the true ability of generalization of casting flaw detection models are evaluated.

3.4.1 Dataset

We collected a novel dataset of aluminium casting radiographs, acquired from various sources, and specifically designed for flaw detection task. Each image was annotated using the Computer Vision Annotation Tool (CVAT) [87] platform. The annotations include precise bounding boxes for each flaw type, along with the grading of each flaw. The labeling process followed the ASTM E2422 reference guidelines [6] to ensure consistency and quality. Both the defect types and their grades were carefully annotated by experts from METU KATAMER. The annotations were reviewed multiple times to improve consistency.

A key distinguishing feature of our dataset is its inclusion of aluminium castings from a wide variety of components and viewpoints. The dataset covers diverse part types, acquired from a defense industry contractor. This part-level diversity introduces significant geometric and structural variability into the dataset, which is often missing in existing casting defect datasets that typically focus on a single component with limited part variety.

Moreover, we include our dataset not only the defects but also flaws that are typically excluded in existing datasets. A flaw is classified as a defect only after exceeding a specific grade threshold, which would result in the material being rejected during non-destructive testing (NDT). For example, a shrinkage cavity may become a defect only after grade 3, while lower-grade shrinkage cavities are considered flaws but not defects. These lower-grade flaws (such as level 1 and level 2) are significantly more subtle and much harder to detect. However, detecting these subtle flaws is critically important because defect grading thresholds may vary across different specifications. For instance, under AMS 2175 Class 1 Grade C [88], a round porosity exceeding grade 3 is classified as a defect, and the corresponding aluminium casting is rejected.

To further emphasize the differences, Figure 2 presents a qualitative comparison between sample images from GDXray, Al-Cast, and our dataset.

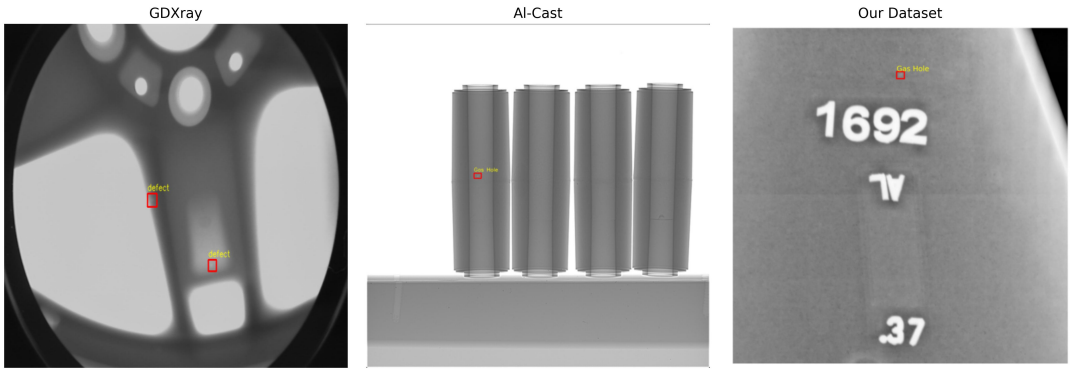


Figure 2: Example images from GDXray, Al-Cast, and our dataset.

In addition, we introduce fine-grained flaw categorization. Rather than grouping all porosity defects together, we explicitly separate elongated porosity and round porosity as distinct classes, following ASTM E2422. Similarly, we differentiate between shrinkage cavities and shrinkage sponge types. This level of fine-grained labeling supports more precise defect characterization, although it significantly increases the difficulty of separating flaw types. Example flaw types from our dataset are illustrated in Figure 3.

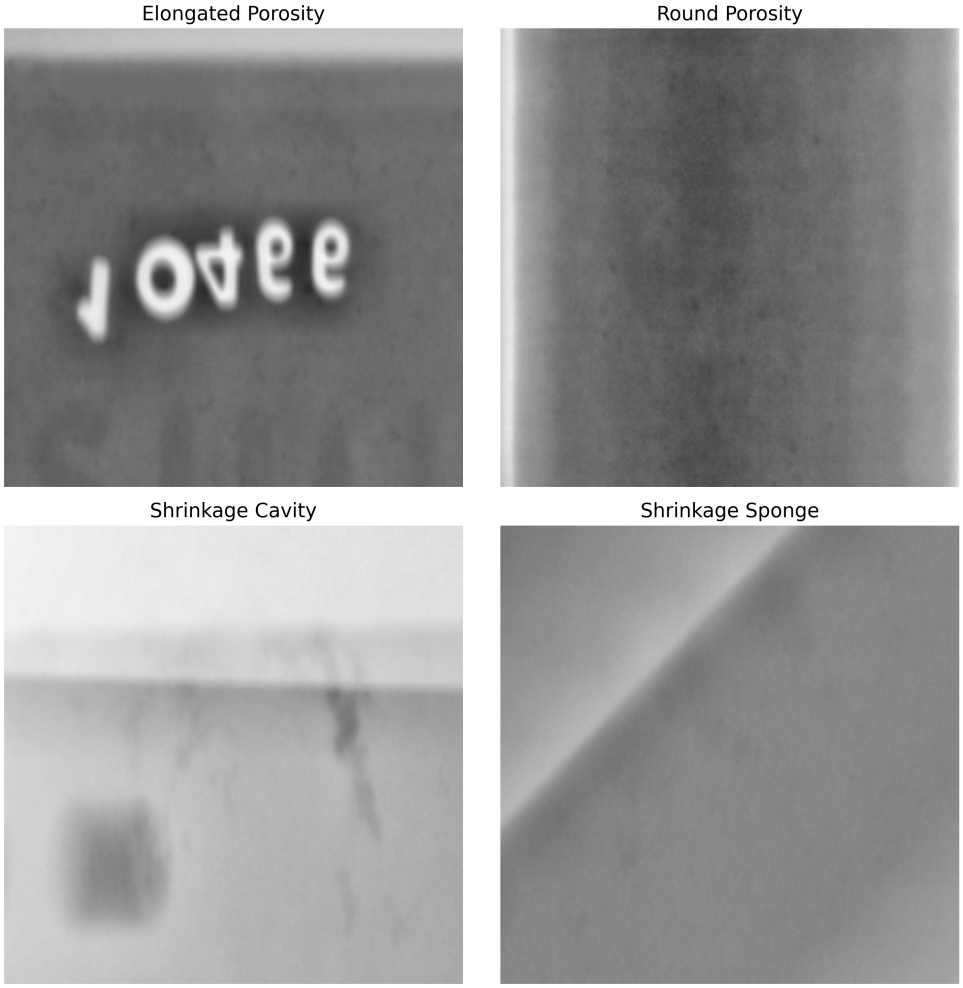


Figure 3: Example fine-grained flaw types from our dataset. The examples include elongated porosity, round porosity, shrinkage cavity, shrinkage sponge. All images are CLAHE enhanced.

As a result, the flaw patterns in our dataset exhibit a broad range of textures, sizes, and contrast levels, providing a more realistic and challenging environment. That is why our dataset offers a stronger benchmark for evaluating the true generalizability of detection models to the casting flaw detection task.

3.4.2 Dataset Statistics

The dataset consists of 856 radiographic images of aluminium castings, with resolutions ranging from 1052×1052 to 3072×3072 pixels. Out of these, 757 images contain annotated flaws, while 99 images are pure negatives with no annotated flaws. The dataset contains a total of 1707 flaw annotations. All the images are saved in the Digital Imaging and Communications in Medicine (DICOM) format and, depending on the imaging device, have a bit depth ranging from 14-bit to 16-bit.

We define **seven classes** in total, following the defect types specified in the ASTM E2422 standard. The defect classes in our dataset include:

- Elongated Porosity
- Round Porosity
- Shrinkage Cavity
- Shrinkage Sponge
- Inclusion (more dense)
- Gas Hole
- Crack

This standard provides reference images for each flaw type across eight distinct grade levels, which correspond to increasing severity and flaw size. Each grade represents a predefined flaw intensity, where grade 1 indicates the smallest acceptable flaw, and grade 8 represents the largest.

Most of the aluminium castings in our dataset fall within the material thickness range covered by this standard, making ASTM E2422 (1/4 inch reference images) directly applicable to our use case. These references are intended to be used in the thickness range up to and including 1/2 inch.

We provide the number of flaw instances per class and the grade distribution across the dataset in Table 1.

Table 1: Distribution of flaw types and grade levels in the dataset.

Class	**Total**	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Elongated Porosity	316	67	121	98	23	6	1	0	0
Round Porosity	623	170	265	133	40	12	3	0	0
Shrinkage Cavity	258	106	61	36	25	13	5	8	4
Shrinkage Sponge	146	37	41	36	20	8	4	0	0
Inclusion (more dense)	150	120	22	6	1	1	0	0	0
Gas Hole	185	154	27	3	1	0	0	0	0
Crack	29	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Total	1707	654	537	312	110	40	13	8	4

It is important to note that although ASTM E2422 does not include crack-type flaws in its reference categories, we have included them in our dataset, as such flaws may also occur in aluminium castings. Since crack-type flaws are not covered by the ASTM E2422 standard, no grade levels are defined for them. In practice, these flaws typically lead to immediate rejection regardless of size or severity.

3.4.3 Baseline Model

For our casting flaw detection task, we benchmarked several object detection models including RetinaNet, FCOS, Faster R-CNN, YOLOv8-n, YOLOv8-s, YOLOv8-m, and YOLOv8-l. We used networks pre-trained on the MS-COCO dataset [66], which is the common adopted practice to transfer general visual knowledge in radiographic NDT domain. Among these, we selected the Faster R-CNN model [29] with a ResNet-50 [89] backbone as the baseline architecture for the rest of this study. Faster R-CNN is a two-stage detector that first generates region proposals via a Region Proposal Network (RPN) and subsequently refines these proposals using a region-based convolutional neural network (RoI head) to perform classification and bounding box regression.

To further strengthen the baseline model, we incorporated several enhancements. First, we explicitly included pure negative samples into the training set. This helps the model better learn the background distribution and reduces false positive detections. Second, we integrated focal loss [26] to both the RPN and RoI heads, but for different purposes. At the RPN stage, focal loss helps the model more effectively differentiate between foreground (potential flaws) and background regions, especially when background anchors dominate the proposals. At the RoI head, focal loss is primarily used to address the class imbalance between the different flaw types and the background, focusing training on harder, misclassified samples and improving robustness against the dominant easy negatives.

The focal loss used in our training is defined as:

$$\mathcal{L}_{\text{focal}} = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{8}$$

where p_t is the predicted probability of the true class label. The hyperparameters α and γ control the weighing of positive and negative samples.

One of the key challenges in our dataset is the presence of extremely small flaws, some of which are as small as 10×10 pixels in high-resolution images. To overcome this, we speculated employing slicing-aided fine-tuning (SF), where the input images are divided into overlapping smaller slices during training combined with full-resolution images to improve the model's sensitivity to small flaws while maintaining detection capability for larger ones. This was followed by slicing-aided inference, where the model predictions on individual slices are aggregated to produce final detections for the full image [90].

3.4.4 Preprocessing

The radiographs in our dataset are single-channel images with varying bit depths of 14-bit and 16-bit, depending on the imaging device. To enhance the visibility of local flaw textures,

and following the findings of [91] [92] [93], we applied Contrast Limited Adaptive Histogram Equalization (CLAHE). Unlike common CLAHE implementations, such as OpenCV's [94] which uses 4096 bins, we specifically implemented a version that uses the number of histogram bins according to the original bit depth of each image to avoid the potential loss of information caused by using a fixed bin size.

Unlike available datasets, we observed that normalizing images to an 8-bit integer range would significantly cause a loss of information for most flaws, some of which becomes purely uniform. That is because the subtle intensity variation of casting defects require higher precision to be effectively represented.

3.4.5 Synthetic Generation of Flaw Radiographs

To generate synthetic radiographs of our flaw classes, we first obtained high-quality reference flaw images directly from the ASTM E2422 standard. Specifically, we cropped reference images for all 8 flaw grades across each defect type provided by the standard. Using these reference images, we applied textual inversion to create custom text embeddings for each defect type separately. Each embedding captures the visual characteristics of a specific flaw type and is associated with a unique pseudo-token that can be used in text prompts to condition the generation process. This enables Stable Diffusion to synthesize new radiographs of specific flaw types guided by the learned embeddings.

We used Stable Diffusion v1.4 as the base generative model throughout this process. Importantly, the flaw concepts we introduce are out-of-distribution with respect to the Stable Diffusion training data. We verified this assumption using CLIP-based text and image retrieval, following a similar methodology utilized in DA-Fusion [73].

3.4.6 Self-Supervised Pretraining with Synthetic Radiographs

To improve the robustness of the backbone used in our flaw detection pipeline, we conducted self-supervised pretraining using these synthetic radiographs. The pretraining dataset was constructed using two sources of synthetic images:

- **Text-to-Image Generated Samples:** For each defect type, synthetic samples were generated via the Stable Diffusion text-to-image pipeline using our custom textual inversion embeddings.

- **Real-to-Synthetic Translated Samples:** To enrich the pretraining data with more realistic global structures, we applied the Stable Diffusion image-to-image translation pipeline. Specifically, we cropped flaw-centered real images from our radiographic dataset and used them as inputs, conditioning the generation on our custom embeddings. To avoid any leakage, image-to-image inputs were taken strictly from the training split.

The aim of text-to-image generation is to introduce object-level variation and encourage the model to learn discriminative features that help separate fine-grained flaw classes. On the other hand, image-to-image translated samples are hypothesized to offer more realistic and coherent radiographic textures. This may help to reduce the domain gap introduced by purely synthetic text-to-image data. However, since casting radiographs lie outside the distribution of natural images used to train Stable Diffusion, the visual quality and semantic consistency of the translated images may be limited.

For pretraining, three settings are evaluated: text-to-image only (T2I), image-to-image only (R2S), and a mixed pool combining both sources. The pretraining was carried out in a multi-positive contrastive learning framework. The model is trained to bring together the representations of different synthetic variants of the same flaw class, while pushing apart the rest. This approach aims to enable the backbone to learn defect-aware features purely from few-shot conditioning using publicly available reference images.

The resulting backbone is later used for downstream tasks such as linear probing and flaw detection on the proposed dataset.

CHAPTER 4

IMPLEMENTATION AND EXPERIMENTS

4.1 Experimental Setup

The synthetic radiograph generation experiments were performed on a workstation equipped with an NVIDIA RTX 4000 ADA GPU, while the rest of the experiments were conducted on workstation equipped with four NVIDIA Quadro RTX 8000 GPUs.

4.2 Radiographic Casting Flaw Detection

4.2.1 Dataset

We divided the dataset into training, validation, and test sets using a stratified sampling strategy. Specifically, 20% of the dataset was reserved as a held-out test set, and the remaining 80% was further split by allocating 20% for validation. The stratification was performed in a class-aware manner to ensure that the distribution of flaw classes is balanced across all splits.

Since most radiographic images in our dataset contain multiple annotated flaws of different types, we assigned each image to the stratified sampling category of its rarest contained flaw class.

4.2.2 Preprocessing

For CLAHE, we used a tile size of 64 and a clip limit of 2 aiming to enhance the local contrast without causing artifacts. As mentioned in Chapter 3.4.4, unlike standard implementations that use a fixed number of histogram bins, we explicitly implemented a version that uses the number of histogram bins to match the bit depth of each image.

The polarity of the images were adjusted to match the polarity convention of ASTM standards, where flaws appear as darker regions against a lighter background.

We normalized the pixel values to a $[0, 1]$ floating-point range based on each image's original bit depth. The single-channel grayscale images were replicated across three channels to create input images compatible with the pretrained backbones. Finally, the normalized images were

further scaled to a $[-1, 1]$ range before being fed into the detection networks, in line with common preprocessing standards.

4.2.3 Training Configuration

All models except the YOLO variants were trained using PyTorch. We used a batch size of 1 due to the high resolution of the images and GPU memory constraints. We also tried gradient accumulation and a shorter-side resize to enable larger batch size, but observed no consistent mAP gains; we therefore report the stable batch size of 1 setup. The training was performed for a total of 200 epochs. Stochastic gradient descent (SGD) with a learning rate of 0.005, momentum of 0.6, and weight decay of 0.0005 was used as an optimizer. Gradient scaling was applied using mixed precision training (FP16) to improve memory efficiency and speed.

We used a learning rate warm-up strategy for the first 10 epochs with a linear increase from 10% of the initial learning rate to the full learning rate. After warm-up, a cosine annealing scheduler was used for the remaining epochs.

For data augmentation, we applied random horizontal and vertical flips, each with a probability of 0.5.

YOLO models, on the other hand, were trained using Ultralytics [95] following the original hyper-parameters.

Model selection was performed based on the best validation achieved during training. Training and evaluation results were tracked using TensorBoard throughout the experiments.

4.2.4 Evaluation Metrics

We evaluated our models using standard COCO object detection metrics. In particular, we focused on mAP at 0.5 IoU threshold (mAP@0.5) and mAR at 100 detections (mAR@100), which are widely used in industrial defect detection.

Each predicted bounding box is classified as one of the following outcomes:

- **True Positive (TP):** A correctly detected flaw where the predicted bounding box sufficiently overlaps with a ground truth box, exceeding the IoU threshold.
- **False Positive (FP):** An incorrect detection where the predicted bounding box either does not sufficiently overlap with any ground truth or detects a non-existent flaw.
- **False Negative (FN):** A ground truth flaw that is not detected by the model.
- **True Negative (TN):** The correct identification of background areas as background.

Precision measures the proportion of correct detections among all predicted bounding boxes:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{9}$$

Higher precision indicates fewer false positive detections.

Recall measures the proportion of correctly detected ground truth flaws:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{10}$$

Higher recall indicates fewer missed flaws.

Intersection over Union (IoU) is the ratio of the overlap area between the predicted and ground truth bounding boxes to their union:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{11}$$

A detection is considered correct if the IoU exceeds a predefined threshold.

Average Precision (AP) summarizes the precision-recall curve for a single class by measuring the area under the curve:

$$\text{AP} = \int_0^1 \text{Precision}(r)\, dr \tag{12}$$

where r denotes recall.

Mean Average Precision (mAP) is the mean of the AP values across all classes:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{13}$$

where N is the number of classes. In this study, we report mean average precision at multiple IoU thresholds ranging from 0.5 to 0.95 with a step size of 0.05, denoted as mAP. We also report mAP at a fixed IoU threshold of 0.5, referred to as mAP50.

Mean Average Recall (mAR) measures the average recall across multiple detection limits and IoU thresholds. It reflects the model's sensitivity to missed detections. Specifically, we use mAR@100, which limits the maximum number of predictions per image to 100.

4.2.5 Benchmark Results

We benchmarked several object detection models to establish a performance baseline on our radiographic casting flaw detection dataset. The evaluated models include RetinaNet, FCOS, Faster R-CNN, YOLOv8-n, YOLOv8-s, YOLOv8-m and YOLOv8-l. The benchmark results are summarized in Table 2.

Table 2: Benchmark results of different object detection models on our dataset.

Model	mAP (0.5:0.95) (%)	mAP50 (%)	mAR@100 (%)
RetinaNet	**11.25**	**32.90**	29.19
FCOS	9.02	29.02	29.13
Faster R-CNN	8.17	25.41	26.58
YOLOv8-n	10.10	26.10	**36.20**
YOLOv8-s	8.61	23.20	27.60
YOLOv8-m	7.95	24.00	28.70
YOLOv8-l	7.02	19.51	28.71

Based on the benchmark results, we selected Faster R-CNN as the baseline model for this study. Faster R-CNN with a ResNet-50 backbone is the standard choice in many self-supervised learning (SSL) evaluations, while providing a strong balance between the detection performance and computational cost in our experiments.

4.2.6 Improving the Baseline Model

For both RPN and ROI heads, we used the following focal loss parameters: $\alpha = 0.25$ and $\gamma = 2.0$, following the original formulation by Lin et al. [26].

For slice-based training and inference using the SAHI framework, we generated the slices according to the following configuration which mostly follows the original implementation:

- **Slice Size:** 512×512
- **Overlap Ratio:** 0.25
- **Detection Confidence Threshold:** 0.3
- **IoU Threshold for Merging:** 0.5

We applied SAHI during both fine-tuning and inference, and full-resolution images were included during training to mitigate potential context fragmentation.

Table 3 summarizes the performance of these configurations using detection metrics. We report overall mAP, mAP50, mAR@100, and scale-specific metrics mAP50s, mAP50m, and mAP50l to better understand performance across flaw sizes for SAHI. The results are given as percentages.

Table 3: Performance comparison of baseline improvements on the test set.

Model Configuration	mAP	mAP50	mAR@100	mAP50s	mAP50m	mAP50l
Faster R-CNN (Baseline)	8.17	25.41	26.58	10.19	9.37	14.44
+ Focal Loss (ROI only)	9.30	27.09	**32.14**	10.87	10.90	**17.75**
+ Focal Loss (RPN + ROI)	**10.89**	**31.55**	31.09	**10.97**	**14.36**	15.39
+ Focal Loss + SF + SAHI	4.70	17.15	14.00	5.94	5.70	3.14

4.3 Synthetic Radiograph Generation

We utilized Stable Diffusion v1.4 to generate synthetic radiographic casting flaw images via textual inversion. For each defect type, we cropped the eight ASTM E2422 reference images, ensuring one image from each grade level, and used these as the training set for textual inversion. Each class was associated with a unique placeholder token (e.g., `<defect-name>`) and trained independently in a few-shot setup.

Unlike the original work [86], which focuses on natural images and employs object-based templates such as "a photo of a clean {}" or style-based templates like "a rendering in the style of {}", we customized the prompt templates to better align with the radiographic domain. Specifically, we only used the following templates during training:

- "a photo of a {}"
- "a cropped photo of the {}"
- "the photo of a {}"
- "a photo of the {}"
- "a variation of a {}"

These templates are more appropriate for the radiographic setting as they emphasize object representation without introducing irrelevant stylistic or descriptive attributes typically used in natural image domains.

To improve the generalization and variance of the learned embeddings, we applied several data augmentations:

- **Random Horizontal Flip:** with probability 0.5
- **Random Vertical Flip:** with probability 0.5
- **Random Rotation:** up to $\pm 45^{\circ}$
- **Gaussian Blur:** applied with probability 0.2
- **Brightness and Contrast Jitter:** applied with probability 0.2

The training configuration is as follows:

- **Learning Rate:** 5×10^{-5}
- **Batch Size:** 2
- **Training Steps:** 13,000
- **Optimizer:** AdamW
- **Learning Rate Scheduler:** Linear warmup for 500 steps, followed by linear decay
- **Resolution:** 512×512
- **Gradient Accumulation:** 1
- **Placeholder Initialization:** the embedding of the token "defect"
- **Mixed Precision:** FP16
- **Loss Function:** Mean Squared Error (MSE) in latent diffusion space between predicted and target noise

The validation configuration is as follows:

- **Validation Frequency:** Every 1,000 steps
- **Validation Set Size:** 30 synthetic samples, generated using the same fixed random seeds
- **Validation Prompt:** a photo of a {}
- **Number of Inference Steps:** 50
- **Guidance Scale:** 7.5

4.3.1 CLIP-Based Retrieval

To verify that our reference radiographs and associated flaw types are out-of-distribution with respect to the Stable Diffusion training data, we perform CLIP-based retrieval experiments on the LAION-400M dataset. While the LAION-5B corpus and its associated k-NN search services are no longer publicly accessible, we use the official PQ64 FAISS indices provided for LAION-400M, which apply a cosine similarity filter threshold of 0.3.

Specifically, we utilize the official image index, text index, and metadata file. This metadata contains 407,314,954 entries, each with a corresponding image URL and caption under the dataset group. All retrieval operations are conducted using OpenAI's CLIP ViT-B/32 model to ensure perfect dimensional alignment with the LAION PQ64 indices.

Text-to-Image Retrieval: For each flaw type, we generate a text prompt of the form:

```
"a radiograph of <defect-name> in metal casting"
```

Each prompt is encoded using CLIP's text encoder, normalized, and used to query the image index for its top-30 most similar image embeddings. The corresponding captions are retrieved from the metadata.

Image-to-Text Retrieval: We also query the text index using CLIP embeddings of our own exemplar reference radiographs that are used to train textual inversion embeddings. Each image is encoded with CLIP's image encoder, normalized, and used to retrieve the top-30 closest text captions.

Full results are presented in Appendix A.1. We removed the duplicate captions.

4.3.2 Evaluation Metrics

Traditional distribution-based metrics such as Fréchet Inception Distance (FID) [58] are widely used for evaluating the quality and diversity of generated images. However, these metrics require a large number of reference images to produce stable and reliable scores. In our setup, since we only have a limited number of ASTM reference images per defect type, these are not suitable.

Instead, we use feature-based methods to focus on two complementary criteria: **quality** and **diversity**.

DINO Score: To assess the visual quality of the generated images, we employed the DINO score, which uses self-supervised features extracted from a pretrained DINO-v2 [96] model. For each image, we extracted a feature vector using the DINO backbone after resizing the image to 224×224 and normalizing to match DINO's expected input.

Let $F_g = \{f_g^1, f_g^2, ..., f_g^n\}$ represent the DINO feature vectors of the n generated images, and $F_r = \{f_r^1, f_r^2, ..., f_r^m\}$ represent the features of the m real reference images. The cosine similarity between each generated image feature and each real image feature is computed as:

$$s_{i,j} = \frac{f_g^i \cdot f_r^j}{\|f_g^i\| \|f_r^j\|} \tag{14}$$

The DINO score is then defined as the mean of pairwise cosine similarities between the generated image and reference images.

Geometric Mean of Standard Deviations: To calculate the diversity among the generated images, we evaluated the spread of DINO feature representations. Specifically, we calculated the standard deviation along each feature dimension and used the geometric mean of these standard deviations as the diversity score [97].

Let $F = \{f_1, f_2, ..., f_N\}$ be the set of DINO feature vectors for N generated images within a class. The standard deviation for the j^{th} feature dimension is denoted as σ_j. The diversity

score D is computed as:

$$D = \left(\prod_{j=1}^{d} \sigma_j \right)^{\frac{1}{d}} \tag{15}$$

where d is the feature vector dimensionality.

4.3.3 Training Configuration Experiments

The convergence of each textual inversion embedding was monitored using 30 validation images generated from the corresponding prompts, evaluated at every 1,000 training steps. To ensure a fair comparison across experiments, all validation steps were performed using the same noise patterns, fixed by the random seed of the noise generator. The guidance scale for every model during validation is also fixed at 7.5.

The effects of varying the number of vectors per embedding and the sample size of reference images were investigated. These experiments were conducted across different random seeds to account for variability in training. The results reported here correspond to the best-performing seed, and best-performing steps for each configuration.

Table 4: Quality and Diversity Metrics for Elongated Porosity

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	0.591	0.638
1	8	**0.613**	0.625
2	5	0.599	0.624
2	8	0.588	0.657
4	5	0.584	0.533
4	8	0.573	0.726

Table 5: Quality and Diversity Metrics for Round Porosity

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	**0.667**	0.558
1	8	0.640	0.563
2	5	0.637	0.630
2	8	0.600	0.617
4	5	0.553	0.675
4	8	0.556	0.696

Table 6: Quality and Diversity Metrics for Shrinkage Cavity

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	0.503	0.637
1	8	**0.562**	0.585
2	5	0.533	0.650
2	8	0.543	0.695
4	5	0.535	0.667
4	8	0.523	0.637

Table 7: Quality and Diversity Metrics for Shrinkage Sponge

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	0.565	0.574
1	8	**0.599**	0.669
2	5	0.488	0.681
2	8	0.520	0.643
4	5	0.504	0.727
4	8	0.500	0.773

Table 8: Quality and Diversity Metrics for Inclusion (More Dense)

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	0.625	0.708
1	8	**0.635**	0.567
2	5	0.609	0.621
2	8	0.627	0.644
4	5	0.616	0.641
4	8	0.588	0.650

Table 9: Quality and Diversity Metrics for Gas Hole

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	5	**0.625**	0.604
1	8	0.621	0.644
2	5	0.577	0.658
2	8	0.607	0.636
4	5	0.591	0.629
4	8	0.610	0.658

Table 10: Quality and Diversity Metrics for Crack

Number of Vectors	Number of Samples	Average DINO Score	Geometric Mean STD
1	1	**0.679**	0.580
2	1	0.625	0.626
4	1	0.592	0.657

4.3.4 Inference Guidance Experiments

To further analyze the effect of the guidance scale during inference, we conducted additional experiments using the best-performing textual inversion models identified in the previous section. Specifically, we varied the **guidance scale** from 2.5 to 15.0, in increments of 2.5.

For each guidance scale, we generated 1,000 synthetic samples per defect type, following similar evaluation methodology described in Section 4.3.3. To ensure a fair comparison across all experiments, we used the same fixed random seeds and the same initial noise patterns throughout the entire generation process while keeping the number of inference steps at 50.

We present the results in the Figure 4, which show the trends of DINO Score and geometric mean standard deviation as a function of the guidance scale.

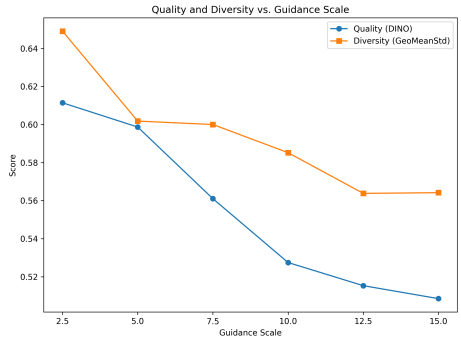
4.4 Self-supervised Pretraining with Synthetic Data

4.4.1 Exploring the Effect of Quality and Diversity

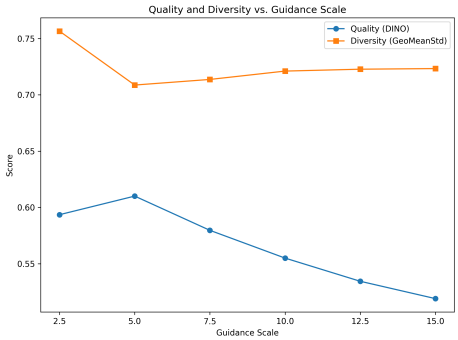
Before constructing the final pretraining datasets, we investigated how synthetic image quality and diversity affect representation learning. As we already generated samples for different guidance scales, two synthetic datasets were formed by selecting guidance scales on a per-class basis as given in Table 11. The downstream results reported in Section 4.4.4.

Table 11: Per-class guidance scales selected for optimizing quality and diversity.

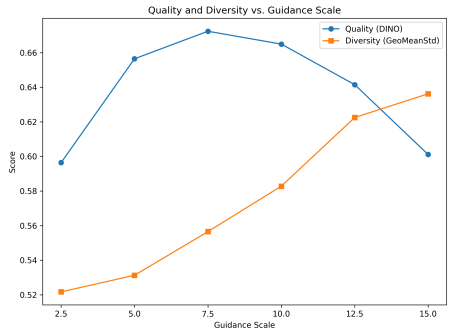
Class	$g^{\star}_{\text{qual}}$	$g^{\star}_{\text{div}}$
Shrinkage Cavity	2.5	2.5
Shrinkage Sponge	5.0	2.5
Elongated Porosity	5.0	15.0
Round Porosity	7.5	15.0
Gas Hole	5.0	2.5
Inclusion	5.0	15.0
Crack	2.5	15.0



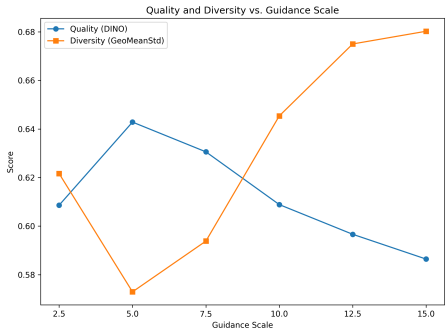
(a) Shrinkage Cavity



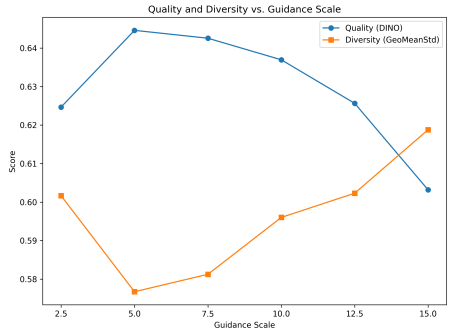
(b) Shrinkage Sponge



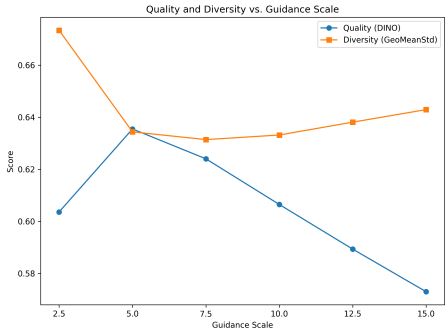
(c) Round Porosity



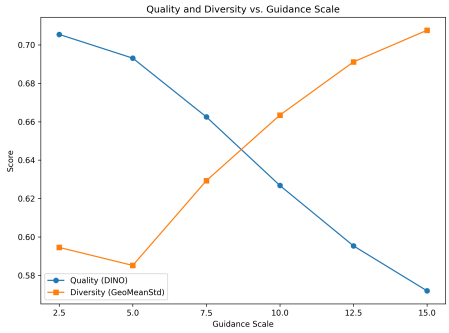
(d) Elongated Porosity



(e) Inclusion (More Dense)



(f) Gas Hole



(g) Crack

Figure 4: Effect of guidance scale on quality and diversity across all flaw types. Each plot shows the relationship between guidance scale, DINO score (quality), and geometric mean of DINO feature standard deviations (diversity).

4.4.2 Pretraining Dataset Construction

We utilized the guidance scales favoring diversity per flaw for generating the pretraining dataset.

Text-to-Image Samples: For each flaw class, we generated 10,000 synthetic samples. In total, T2I dataset consist of 70,000 images.

Real-to-Synthetic Samples: The generation process operates directly on 512×512 crops centered around annotated flaws from the training split of our dataset. The original images are in mixed 14- and 16-bit format, and the pipeline was configured to process these without loss of detail. In total, R2S dataset consist of 68,400 images.

For each class, its own class-specific textual inversion embedding was used to guide the generation. The following parameters were applied throughout:

- **Strength:** 0.1
- **Prompt Template:** "a photo of a {name}"
- **Number of Synthetic Images per Sample:** 50

The final pretraining datasets were constructed separately for T2I, R2S, and mixed variants across all flaw classes. Example images from both T2I and R2S sets for round porosity can be seen in Figure 5.

4.4.3 Pretraining Configuration

For self-supervised pretraining, we modified the StableRep [76] framework with a multi-positive contrastive learning objective. Instead of the original Vision Transformer implementation of StableRep, We used a ResNet-50 backbone trained from scratch (i.e., without ImageNet initialization) for fairness.

The encoder architecture consists of a standard ResNet-50 followed by a 3-layer Multi-layer Perceptron (MLP) projection head. The MLP maps 2048-dimensional visual features to a 128-dimensional embedding space via a hidden layer of size 2048:

- **Backbone:** ResNet-50
- **Projection Head:** Linear→BN→ReLU→Linear→BN→ReLU→Linear
- **Hidden Dim:** 2048, **Embedding Dim:** 128

The model was trained using AdamW with the following optimization parameters:

- **Optimizer:** AdamW

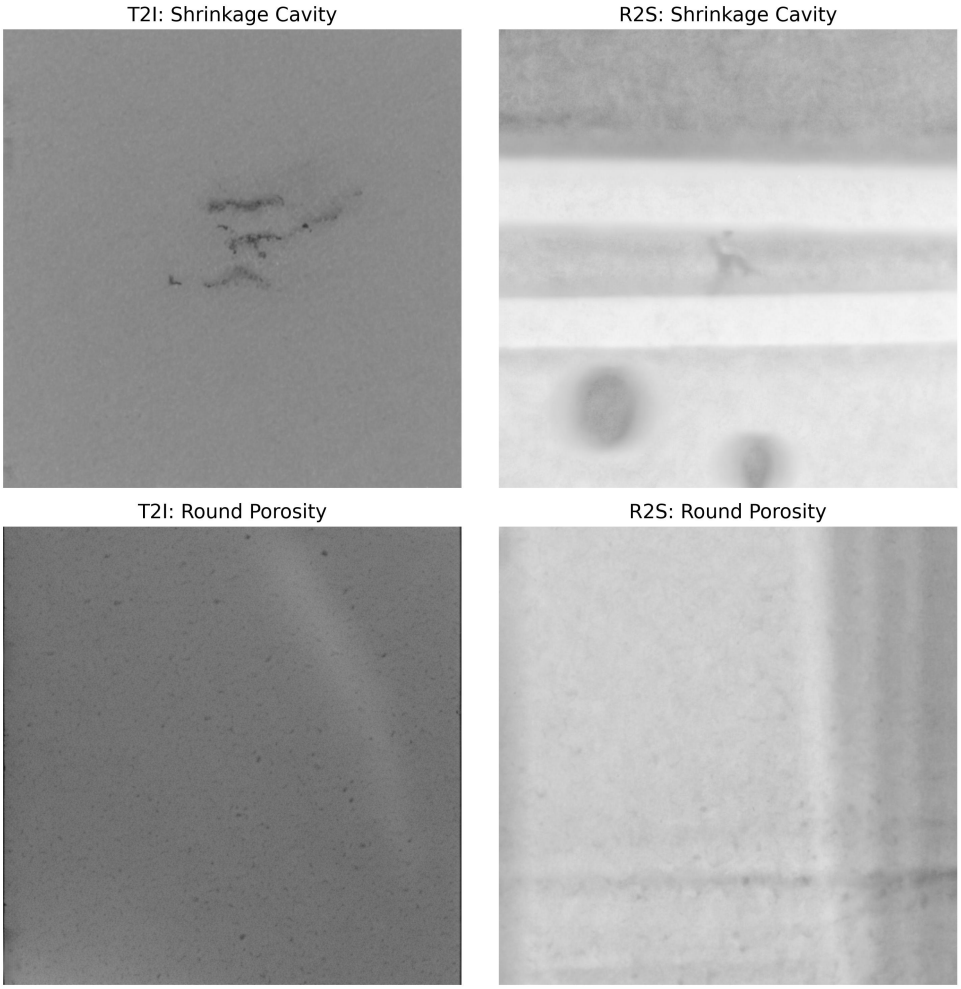


Figure 5: Example synthetic samples for the *Round Porosity* class. Left: Text-to-image generations. Right: Real-to-synthetic translations using image-to-image pipeline.

- **Learning Rate:** 4e-3
- **Weight Decay:** 0.1, **Betas:** (0.9, 0.95)
- **Warmup Epochs:** 5.0
- **Total Epochs:** 400

The training was done on four GPUs, with a batch size of 32 per GPU and four positives per sample.

To better simulate the variety of radiographic conditions, the augmentation pipeline was modified as:

- Resize to 224, color jitter (brightness, contrast), Gaussian blur, random flips, and randomly downsample to {64, 128, 192} resolutions.

Unlike SimCLR and StableRep, we avoid applying local or random crops during pretraining. This is because the flaws in our dataset such as inclusion, gas hole, and crack often occupy a small region within the image. Applying random cropping risks losing critical structural details or generating background-only views that provide no meaningful learning signal.

All experiments were run with mixed-precision training. We used the temperature-scaled multi-positive contrastive loss, with cosine-scheduled temperature decay between 0.3 and 0.02.

4.4.4 Downstream Task Evaluation

We follow the evaluation by the only prior work in self-supervised learning for casting defects [7]. The pretrained backbone is frozen and evaluated via linear probing, where a supervised classifier is trained on cropped bounding boxes from our own dataset to distinguish between fine-grained flaw types due to the lack of any public benchmark classification dataset in this domain. The classifier is a simple linear head trained on top of the 2048-dimensional output of the ResNet-50 backbone. We train multiple linear classifiers using different base learning rates (0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.5) with batch normalization. Training is performed using SGD with momentum 0.6, over 100 epochs with a cosine learning rate schedule. Each classifier is trained using a batch size of 64. The best-performing one is selected based on top-1 accuracy. The maximum top-1 accuracy reported as the evaluation metric, which is computed every 10 epochs.

We first evaluate the impact of synthetic generation strategy on SSL pretraining using the quality-optimized and diversity-optimized datasets introduced in Sec. 4.4.1. Linear probing accuracy is reported in Table 12.

Table 12: The effect of synthetic sample quality and diversity for SSL pretraining.

Pretraining Dataset	**Top-1 Accuracy (%)**
Quality-Optimized	70.21
Diversity-Optimized	**71.09**

We then evaluate three configurations of our pretraining dataset (introduced in Sec. 4.4.2) for SSL pretraining:

- **T2I**: text-to-image samples only.
- **R2S**: real-to-synthetic translated samples only.
- **Mixed**: combination of T2I and R2S samples.

Table 13 summarizes the results, comparing our SSL-pretrained backbones against ImageNet-pretrained and randomly initialized ResNet-50 models on the flaw classification task.

Table 13: Linear probing results on the flaw classification task. The SSL-pretrained ResNet-50s, ImageNet-initialized and randomly initialized baselines.

Backbone	**Top-1 Accuracy (%)**
Random Init	61.06
ImageNet-pretrained	71.68
SSL-pretrained (R2S)	70.80
SSL-pretrained (T2I)	**74.63**
SSL-pretrained (Mixed)	71.09

End-to-end Finetuning for Detection. We also integrate the pretrained backbone into the Faster R-CNN model and fine-tune it on our full flaw detection dataset similar to [7]. The Faster R-CNN detector is initialized with a ResNet-50 backbone and Feature Pyramid Network (FPN), where the backbone weights are transferred from our self-supervised pretraining stage. The Region Proposal Network (RPN) and detection heads are initialized from scratch.

The detector is fine-tuned with a multi-step learning rate schedule over 300 epochs to facilitate convergence of SSL-pretrained backbones. To better measure feature robustness, the augmentation policy is strengthened by adding random rotations and color jitter (brightness and contrast); all other settings remain identical to the baseline. Evaluation is conducted similarly using mAP, mAP50, and mAR@100, following standard COCO-style metrics. Table 14 compares ImageNet-pretrained backbones with SSL-pretrained backbones trained on different synthetic datasets.

Table 14: Faster R-CNN detection results on our flaw detection dataset. Comparison of backbone initialization strategies in terms of overall mAP, mAP50, and mAR@100.

Backbone Initialization	**mAP (%)**	**mAP50 (%)**	**mAR@100 (%)**
ImageNet-pretrained ResNet-50	6.94	20.53	20.80
SSL-pretrained ResNet-50 (R2S)	4.53	14.34	16.74
SSL-pretrained ResNet-50 (T2I)	**7.05**	**23.39**	**21.14**
SSL-pretrained ResNet-50 (Mixed)	6.49	19.82	17.94

CHAPTER 5

DISCUSSION

5.1 Analysis of Detection Performance and Impact of the Dataset

The benchmark results of object detection models on our dataset in Table 2 showed significantly lower mAP and mAR values compared to results reported on public datasets, namely GDXray and Al-Cast. This performance gap can be attributed to the inherent challenges of our dataset.

Unlike GDXray, which primarily focuses on binary defect detection (defect or no defect), our dataset introduces fine-grained flaw categories that require the models to differentiate between subtle visual patterns. Additionally, we included all ASTM grade levels for each flaw type, leading to the inclusion of very subtle flaws that are difficult to detect even for human experts. This is also why our dataset contains radiographs with higher bit depth than those in either GDXray or Al-Cast. In contrast, the GDXray dataset includes more visually pronounced defects, which simplifies the detection task.

Moreover, although GDXray provides 2,727 casting radiographs, they are organized into 67 series. Each series often consists of rotated or cropped versions of the same material imaged under different views. While these views are physically acquired (rather than synthetically augmented), the dataset primarily includes a limited range of parts, namely aluminum wheels and knuckles. Many private industrial datasets also share this limitation. They often focus on a specific part or fixed geometry, which causes lack of diversity in both flaw types and part geometries.

The Al-Cast dataset, although offering two types of defect classes, contains only 187 original images. Therefore, Al-Cast heavily relies on geometric augmentations such as rotations to increase the dataset size synthetically. The limited number of samples and constrained diversity in part geometry and viewpoints in Al-Cast likely leads to overfitting and inflated performance metrics that do not generalize well to more realistic scenarios in NDT.

In contrast, our dataset:

- Specifically annotated using ASTM E2422 standard.
- Contains flaws that are more subtle due to the presence of all grade levels.
- Includes variety of parts with different geometric configurations.

- Fine-grained defect categories are explicitly annotated.

These factors increase the difficulty of the detection task and contribute to the lower performance observed across all evaluated models. However, this difficulty more accurately reflects the real-world challenges encountered in industrial radiographic inspection, where distinguishing subtle flaws under diverse conditions is essential. A reliable evaluation of model generalizability in NDT requires such diversity in part geometry, perspective, and fine-grained defect separation.

Benchmark Results. Looking at Table 2, RetinaNet achieved the highest mAP score of 11.25% and the best mAP50 score of 32.90%, indicating strong overall detection precision across varying IoU thresholds. This can be attributed to its use of focal loss, which is particularly effective in handling class imbalance. This is useful in our dataset where certain flaw categories such as crack and inclusion are underrepresented (see Table 1).

FCOS and YOLOv8-n also performed competitively, achieving mAP scores of 9.02% and 10.10%, respectively. Notably, YOLOv8-n recorded the highest recall (mAR@100) at 36.20%, suggesting a stronger ability to localize more flaws, however, with slightly reduced precision.

Interestingly, larger YOLOv8 variants underperformed relative to their smaller counterparts. This could be due to overfitting or insufficient data to fully exploit their higher capacity, highlighting the importance of model size selection in low-data regimes.

Faster R-CNN did not achieve the absolute best performance across all metrics, but it provided a competitive balance and is widely used in self-supervised learning evaluations. Due to its two-stage detection architecture and ResNet-50 backbone, it offers a robust framework for future experiments and serves as a reliable baseline for the rest of this study.

5.2 Improving the Baseline Model

The integration of focal loss for both RPN and ROI heads resulted in a measurable improvement in detection performance, as summarized in Table 3. When focal loss was applied solely to the ROI head, the model achieved an improvement over the baseline Faster R-CNN, increasing the overall mAP from 8.17 % to 9.30 %, and mAP50 from 25.41 % to 27.09 %. Moreover, mAR@100 showed a more substantial gain from 26.58 % to 32.14 %.

The combined application of focal loss to both RPN and ROI stages further amplified these gains, boosting mAP to 10.89 % and mAP50 to 31.55 %. Interestingly, the recall slightly decreased to 31.09 %, suggesting a trade-off between precision-focused improvements in region proposal quality and overall retrieval. Still, this configuration showed the strongest gains.

In the RPN, focal loss effectively suppressed the overwhelming influence of easy background examples by down-weighting well-classified negatives, which is particularly relevant in radiographic flaw detection where flaws can be sparse relative to the image size. In the ROI head, focal loss further helped address class imbalance among flaw types. The combination

of these changes improved the model's sensitivity to rare or low-contrast defects while maintaining robustness against false positives.

We also experimented with slice-aided fine-tuning and inference using SAHI. However, this approach led to worse performance. One likely reason is that casting flaws particularly elongated and round porosity are spatially extensive and often spread over a broad area. When such regions are divided into slices, the model is deprived of important global context necessary for accurately identifying the defect as a whole. Moreover, slicing may increase the number of false positives in our setting. Certain natural variations in the casting material, such as minor density fluctuations, surface marks, or grain patterns, can resemble flaws when viewed in isolation. Without the broader structural context, these features are more likely to be misclassified as flaws.

Another observation is that SAHI's assumption of a relatively consistent object-to-image size ratio, as seen in datasets like COCO, does not hold in our domain. Flaw sizes can vary drastically, from tiny gas holes to extensive areas of porosity. This makes fixed slicing configurations suboptimal. In our experiments, even after including full-resolution images during training to mitigate fragmentation issues, SAHI still underperformed.

5.3 Analysis of Synthetic Radiograph Generation

The results of our synthetic generation experiments showed several important insights regarding the effect of key configuration choices on the quality and diversity of generated flaw images.

CLIP-based Retrieval. In both retrieval directions, none of the top-30 neighbors reference terms such as "casting" or any of the defect names used in our work. The cosine similarities are very low and text queries are dominated by medical X-rays while image queries bring unrelated concepts. These results provide strong empirical evidence that our images and associated prompts were out-of-domain for the CLIP, and by extension, for Stable Diffusion's training data. This supports the claim that our synthetic data and learned embeddings do not overlap with the original diffusion model training data.

Number of Vectors and Reference Samples. We evaluated three vector sizes (1, 2, and 4) across different numbers of reference samples (5 and 8), with the results summarized in Tables 4–10. Contrary to the expectation that increased embedding capacity leads to richer representations, we observed that using a single learned vector yielded the highest DINO scores across most flaw types. Increasing the number of vectors to two slightly reduced quality, and the degradation became more pronounced with four vectors, especially for the cases with 8 reference images. This suggests that over-parameterization may harm performance when the number of reference images is limited, possibly due to overfitting or unstable token learning.

We hypothesize that this trend may be attributed to the nature of radiographic casting defects themselves. Unlike natural objects with complex semantics and varying spatial configurations, casting flaws are typically characterized by localized, low-level texture variations. These features do not require multiple token embeddings to represent diverse high-level concepts, making a single vector both sufficient and more stable.

Across all flaw types, using one vector was selected as the optimal configuration as it consistently yielded the best DINO scores. Moreover, increasing the number of reference images from 5 to 8 also improved quality, showing the role of reference diversity in learning robust embeddings. An exception was the "Gas Hole" and "Round Porosity" category, where additional samples did not yield better performance. We believe this is potentially due to lower intra-class variability of the reference images for these flaws.

Although we report both quality (DINO score) and diversity (geometric mean of DINO feature standard deviations) for completeness, we selected the best-performing models solely based on DINO score. This decision was made because diversity estimates derived from only 30 validation samples are statistically unreliable. A more robust analysis of diversity is presented later in Section 4.3.4, using 1,000 generated samples per guidance scale setting. Therefore, model selection at this stage focuses strictly on quality.

Effect of Guidance Scale. Varying the guidance scale during inference had a notable influence on generation characteristics. As shown in Figure 4, increasing the guidance scale typically improved DINO scores up to a point (often peaking around 5.0), after which performance started to decline. This is consistent with the expected behavior of classifier-free guidance: higher values encourage stronger conditioning but may lead to over-saturation of learned tokens. These results aligns with MixDiff [79] which also reports that higher guidance scales reduce image quality, with the best quality at low-mid guidance reported on ImageNet.

Increasing classifier-free guidance too much can introduce high-frequency artifacts. In our domain, where flaw classes have intrinsically low semantic variability (mostly size, shape, and contrast), the observed rise in the diversity proxy at large guidance is thus interpreted as artifact-driven feature spread rather than expanded semantic coverage.

A recurring pattern observed across nearly all defect classes was a trade-off between quality and diversity. Configurations that show higher DINO scores (suggesting greater fidelity to real images) often exhibited slightly lower diversity, and vice versa.

Defect-Specific Patterns. The sensitivity to guidance scale and embedding configuration varied across defect types. For example, defects like *shrinkage cavity* and *crack* showed higher DINO scores at lower guidance scales, indicating that their visual structure is easier to learn and replicate even with weak conditioning.

Stability and Reproducibility. Fixing the random seed and latent noise across runs allowed for a more controlled comparison, revealing subtle yet reproducible trends in performance.

This design choice ensured that improvements were not artifacts of favorable sampling and helped isolate the effects of each configuration.

Overall, these experiments demonstrate that tuning textual inversion setups in the radiographic domain requires balancing between embedding capacity, reference diversity, and inference-time guidance. Importantly, optimal configurations can differ depending on the defect type, suggesting that per-class customization is beneficial.

5.4 Analysis of Self-supervised Pretraining with Synthetic Data

Our experiments demonstrate that self-supervised pretraining on synthetic data can lead to strong visual representations for downstream casting NDT tasks that include both coarse and fine-grained flaw categories. With only synthetic data, the SSL-pretrained ResNet-50 backbone outperforms ImageNet-pretrained counterparts on downstream tasks.

Quality–Diversity Ablation. Despite lower image quality, the diversity-favoring sets yielded better linear-probe accuracy (e.g., 71.09% vs. 70.21% shown in Table 12). This matches reports that, in synthetic-only regimes, lower-quality generations can act as stronger augmentations and benefit SSL, even when "diversity" stems from artifact-sensitive variance [79, 76].

Linear Probing. In the linear probing results (Table 13), T2I self-supervised pretraining attains the highest top-1 accuracy at **74.63%**, outperforming both the ImageNet-pretrained baseline (**71.68%**) and the randomly initialized model (**61.06%**). The R2S setting yields **70.80%** which is comparable but slightly below ImageNet, while the Mixed configuration reaches **71.09%**. These trends suggest that text-to-image synthetic views provide the most class-discriminative, defect-aware features for flaw classification.

Detection Fine-tuning. On full-image detection (Table 14), T2I again performs best with **mAP=7.05%**, **mAP50=23.39%**, and **mAR@100=21.14%**. Relative to ImageNet (6.94% mAP, 20.53% mAP50, 20.80% mAR@100), improvements are concentrated at $\mathrm{IoU} = 0.5$ ($+2.86$ points, $\sim 14\%$ relative) with a modest recall gain ($+0.34$ points). This pattern is consistent with our multi-positive contrastive pretraining, which sharpens inter-class separation (precision) more than it improves region/foreground separation (recall). In contrast, R2S underperforms significantly, and Mixed lags ImageNet.

R2S Configuration. We hypothesize that the poor R2S results stem from low visual quality of real-to-synthetic samples, which arises from a combination of domain mismatch and limits of textual inversion training. Specifically, casting X-ray images fall significantly outside the distribution of Stable Diffusion's original training data, which biases the model toward natural scene priors. Moreover, our textual inversion tokens were trained on uniform-background defect crops, likely encouraging the model to overwrite contextual material structures during

image-to-image translation. This causes R2S outputs to be visually degraded and semantically inconsistent, limiting their utility in SSL pretraining.

Consistently, R2S achieves linear-probe accuracies comparable to ImageNet on tighter and more centered crops but lags significantly in detection, where sensitivity to background texture and global context is higher.

Mixed Configuration. Treating T2I and R2S views as positives for the same instance forces invariance across two visually inconsistent domains. This can blur decision boundaries and reduce class margins, explaining why Mixed trails T2I in both linear probe and detection.

Overall, our findings validate that carefully curated and guided T2I synthetic pretraining can serve as a viable foundation for learning domain-specific representations in a fully self-supervised manner.

CHAPTER 6

CONCLUSION AND FUTURE WORK

This work addressed a critical challenge in non-destructive testing (NDT) for industrial quality assurance: the automated detection of casting flaws in aluminum components using radiographic images. While X-ray imaging is widely used in manufacturing, the lack of publicly available, high-quality, and diverse annotated datasets has severely limited the development and evaluation of deep learning solutions in this domain.

To address this, we introduced a new radiographic flaw detection dataset specifically curated for the casting domain. This dataset includes a wide spectrum of flaw grades that are often overlooked in other benchmarks, as well as both coarse and fine-grained flaw categories such as different porosity and shrinkage types, cracks, inclusions, and gas holes. The annotations follow industrial standards (ASTM) and include high-resolution and high bit-depth radiographs, offering a realistic, diverse, and representative benchmark for evaluating NDT systems. By providing this dataset, we aim to accelerate research and development in industrial inspection by offering a benchmark that reflects the practical challenges of flaw detection in real manufacturing settings.

We systematically evaluated detection performance on this dataset using several object detection models. Moreover, we further analyzed the impact of loss function choices such as focal loss. Our evaluation revealed key limitations of approaches like slicing aided methods when applied to flaw detection, highlighting the importance of preserving global defect context.

In addition to our detection evaluations, we proposed a synthetic data generation framework tailored for radiographic inspection tasks. The flaw-specific embeddings used in this framework were derived from just a few publicly available reference images, ensuring no data leakage or dataset-specific bias. This also allows the embeddings to be reused across different datasets in future studies.

The generated synthetic dataset was used for self-supervised pretraining. We demonstrated that self-supervised learning can effectively benefit from synthetic images generated from out-of-distribution concepts—specifically, flaw types introduced solely through textual inversion of a few standard reference images. Our SSL-pretrained model achieved the best performance in both classification and detection tasks, surpassing ImageNet-pretrained and randomly initialized baselines. These results validate the effectiveness of domain-specific synthetic pretraining for casting flaws in the NDT applications.

In conclusion, this work contributes to the NDT community not only by introducing a challenging and realistic flaw detection dataset, but also by demonstrating a synthetic pretraining pipeline that enables robust representations of flaws without requiring large amounts of real data.

Future Work and Limitations

While our work introduces a practical and extensible foundation for automated casting flaw detection, several directions remain open for future exploration. Future work could explore integrating severity grading either as a regression task or as an ordinal classification problem. This would enable models to distinguish not only the type but also the extent of flaws. This is a critical need in industrial settings where the decision to reject or accept a part depends heavily on flaw severity.

Moreover, the performance of the detection models are still low reflecting the inherent difficulty of the dataset. Architectural enhancements and curriculum learning strategies could improve results further.

Another limitation arises from the absence of publicly available datasets that include similar fine-grained defect classes. This made it impossible to compare the robustness and generalizability of our learned representations across different datasets. To address this, we conducted consistent and well-defined experiments on our own dataset to establish reliable baseline results for the literature, which future works can use for comparison and benchmarking.

One other promising direction is the principled selection of synthetic images prior to self-supervised pretraining. Rather than using the entire set of generated samples, future work could investigate how focusing the training data around task-relevant features or distributions might enhance representation learning. Careful selection can reduce domain shift to yield stronger, more stable improvements in downstream tasks.

Lastly, due to resource limitations, we did not exhaustively explore SSL variations, scaling laws, or semi-supervised extensions. These aspects could further inform the design of domain-specific SSL pipelines for NDT.

REFERENCES

[1] C. Bharambe, M. Jaybhaye, A. Dalmiya, C. Daund, and D. Shinde, "Analyzing casting defects in high-pressure die casting industrial case study," *Materials Today: Proceedings*, vol. 72, pp. 1079–1083, 2023.

[2] M. V. Glazoff, V. S. Zolotorevsky, and N. A. Belov, *Casting aluminum alloys*. Elsevier, 2010.

[3] S. Cumblidge, A. D'Agostino, S. Morrow, C. Franklin, and N. Hughes, "Review of human factors research in nondestructive examination," in *US Nuclear Regulatory Commission-Pacific Northwest National Laboratory) 7th European-American Workshop on Reliability of NDE*, 2017.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[5] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

[6] ASTM International, "Standard reference radiographs for inspection of aluminum and magnesium castings," 2022. ASTM E2422-22.

[7] E. Intxausti, D. Skočaj, C. Cernuda, and E. Zugasti, "A methodology for advanced manufacturing defect detection through self-supervised learning on x-ray images," *Applied Sciences*, vol. 14, no. 7, 2024.

[8] S. Hernández, D. Sáez, and D. Mery, "Neuro-fuzzy method for automated defect detection in aluminium castings," in *Image Analysis and Recognition: International Conference, ICIAR 2004, Porto, Portugal, September 29-October 1, 2004, Proceedings, Part II 1*, pp. 826–833, Springer, 2004.

[9] X. Li, S. K. Tso, X.-P. Guan, and Q. Huang, "Improving automatic detection of defects in castings by applying wavelet technique," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 6, pp. 1927–1934, 2006.

[10] J. Zhang, Z. Guo, T. Jiao, and M. Wang, "Defect detection of aluminum alloy wheels in radiography images using adaptive threshold and morphological reconstruction," *Applied Sciences*, vol. 8, no. 12, 2018.

[11] J. Zhang, L. Hao, T. Jiao, L. Que, and M. Wang, "Mathematical morphology approach to internal defect analysis of a356 aluminum alloy wheel hubs," *AIMS Mathematics*, vol. 5, no. 4, pp. 3256–3273, 2020.

[12] R. Cogranne and F. Retraint, "Statistical detection of defects in radiographic images using an adaptive parametric model," *Signal Processing*, vol. 96, pp. 173–189, 2014.

[13] C. Arteta, "Automatic defect recognition in x-ray testing using computer vision," pp. 1026–1035, 03 2017.

[14] B. Wu, J. Zhou, X. Ji, Y. Yin, and X. Shen, "Research on approaches for computer aided detection of casting defects in x-ray images with feature engineering and machine learning," *Procedia Manufacturing*, vol. 37, pp. 394–401, 2019. Physical and Numerical Simulation of Materials Processing IX.

[15] W. Du, H. Shen, J. Fu, G. Zhang, and Q. He, "Approaches for improvement of the x-ray image defect detection of automobile casting aluminum parts based on deep learning," *NDT E International*, vol. 107, p. 102144, 2019.

[16] D. Wangzhe, H. Shen, F. Jianzhong, G. Zhang, X. Shi, and Q. He, "Automated detection of defects with low semantic information in x-ray images based on deep learning," *Journal of Intelligent Manufacturing*, vol. 32, 01 2021.

[17] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Detnet: A backbone network for object detection," *arXiv preprint arXiv:1804.06215*, 2018.

[18] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, 2018.

[19] L. Duan, K. Yang, and L. Ruan, "Research on automatic recognition of casting defects based on deep learning," *IEEE Access*, vol. 9, pp. 12209–12216, 2021.

[20] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[21] L. Jiang, Y. Wang, Z. Tang, Y. Miao, and S. Chen, "Casting defect detection in x-ray images using convolutional neural networks and attention-guided data augmentation," *Measurement*, vol. 170, p. 108736, 2021.

[22] L. Xue, J. Hei, Y. Wang, Q. Li, Y. Lu, and W. Liu, "A high efficiency deep learning method for the x-ray image defect detection of casting parts," *Measurement Science and Technology*, vol. 33, no. 9, p. 095015, 2022.

[23] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781–10790, 2020.

[24] A. García Pérez, M. J. Gómez Silva, and A. De La Escalera Hueso, "Automated defect recognition of castings defects using neural networks," *Journal of nondestructive evaluation*, vol. 41, no. 1, p. 11, 2022.

[25] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "Gdxray: The database of x-ray images for nondestructive testing," *Journal of Nondestructive Evaluation*, vol. 34, no. 4, p. 42, 2015.

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[27] S. Cheng, J. Lu, M. Yang, S. Zhang, Y. Xu, D. Zhang, and H. Wang, "Wheel hub defect detection based on the ds-cascade rcnn," *Measurement*, vol. 206, p. 112208, 2023.

[28] Z. Cai and N. Vasconcelos, "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162, 2018.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," 2016.

[30] I. E. Parlak and E. Emel, "Deep learning-based detection of aluminum casting defects and their types," *Engineering Applications of Artificial Intelligence*, vol. 118, p. 105636, 2023.

[31] Ultralytics, "YOLOv5: A state-of-the-art real-time object detection system." `https://docs.ultralytics.com`, 2021. Accessed: 2025.

[32] Y. Wang, F. Zuo, S. Zhang, and Z. Zhao, "Progressive frequency-guided depth model with adaptive preprocessing for casting defect detection," *Machines*, vol. 12, no. 3, 2024.

[33] R. Varghese and S. M., "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pp. 1–6, 2024.

[34] C. Hai, Y. Wu, H. Zhang, F. Meng, D. Tan, and M. Yang, "Approach for automatic defect detection in aluminum casting x-ray images using deep learning and gain-adaptive multi-scale retinex," *Journal of Nondestructive Evaluation*, vol. 43, no. 1, p. 29, 2024.

[35] S. Zhang, H. Li, P. Ren, T. Peng, and X. Meng, "A detection method for small casting defects based on bidirectional feature extraction," *Scientific Reports*, vol. 15, no. 1, p. 6362, 2025.

[36] D.-J. Cheng, S. Wang, H.-B. Zhang, and Z.-Y. Sun, "A novel framework for low-contrast and random multi-scale blade casting defect detection by an adaptive global dynamic detection transformer," *Computers in Industry*, vol. 162, p. 104138, 2024.

[37] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," 2020.

[38] H.-B. Zhang, C.-Y. Zhang, D.-J. Cheng, K.-L. Zhou, and Z.-Y. Sun, "Detection transformer with multi-scale fusion attention mechanism for aero-engine turbine blade cast defect detection considering comprehensive features," *Sensors*, vol. 24, no. 5, 2024.

[39] L. Zhang, S.-f. Yan, J. Hong, Q. Xie, F. Zhou, and S.-l. Ran, "An improved defect recognition framework for casting based on detr algorithm," *Journal of Iron and Steel Research International*, vol. 30, no. 5, pp. 949–959, 2023.

[40] L. Li, M. Gao, X. Tian, C. Wang, and J. Yu, "Surface defect detection method of aluminium alloy castings based on data enhancement and crt-detr," *IET Image Processing*, vol. 18, no. 13, pp. 4275–4286, 2024.

[41] Q. Wang, S. Gao, L. Xiong, A. Liang, K. Jiang, and W. Zhang, "A casting surface dataset and benchmark for subtle and confusable defect detection in complex contexts," *IEEE Sensors Journal*, 2024.

[42] J. Xu, X. Hu, Y. Zhang, and Z. Li, "Bhe-yolo: Effective small target detector for aluminum surface defect detection," *Advanced Theory and Simulations*, vol. 7, no. 1, p. 2300563, 2024.

[43] H. Pan, H. Zhao, X. Wei, D. Zhang, B. Dong, and J. Lan, "Study on defect detection of metal castings based on supervised enhancement and attention distillation," *Machine Vision and Applications*, vol. 35, no. 3, p. 57, 2024.

[44] B. Hena, G. Ramos, C. Ibarra-Castanedo, and X. Maldague, "Automated defect detection through flaw grading in non-destructive testing digital x-ray radiography," *NDT*, vol. 2, no. 4, pp. 378–391, 2024.

[45] İsmail Enes Parlak and E. Emel, "Deep learning-based detection of internal defect types and their grades in high-pressure aluminum castings," *Measurement*, vol. 242, p. 116119, 2025.

[46] D. Mery, D. Hahn, and N. Hitschfeld, "Simulation of defects in aluminium castings using cad models of flaws and real x-ray images," *Insight-Non-Destructive Testing and Condition Monitoring*, vol. 47, no. 10, pp. 618–624, 2005.

[47] D. Mery, "Aluminum casting inspection using deep object detection methods and simulated ellipsoidal defects," *Machine Vision and Applications*, vol. 32, no. 3, p. 72, 2021.

[48] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg, "Quantifying the effect of x-ray scattering for data generation in real-time defect detection," *Journal of X-Ray Science and Technology: Clinical Applications of Diagnosis and Therapeutics*, vol. 32, p. 1099–1119, July 2024.

[49] S. Bosse, D. Lehmhus, and S. Kumar, "Automated porosity characterization for aluminum die casting materials using x-ray radiography, synthetic x-ray data augmentation by simulation, and machine learning," *Sensors*, vol. 24, no. 9, 2024.

[50] B. Gosswami, R. Chanda, and T. Wittenberg, "Synthetic x-ray image generation for non-destructive testing using generative adversarial networks," *e-Journal of Nondestructive Testing*, 2024.

[51] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[52] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[53] A. García-Pérez, M. Gómez-Silva, and A. de la Escalera-Hueso, "Improving automatic defect recognition on gdxray castings dataset by introducing genai synthetic training data," *NDT E International*, vol. 151, p. 103303, 2025.

[54] R. Guo, H. Liu, G. Xie, and Y. Zhang, "Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10844–10853, 2021.

[55] L. Li, P. Wang, J. Ren, Z. Lü, X. Li, H. Gao, and R. Di, "Synthetic data augmentation for high-resolution x-ray welding defect detection and classification based on a small number of real samples," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108379, 2024.

[56] X. Wang, F. He, and X. Huang, "A new method for deep learning detection of defects in x-ray images of pressure vessel welds," *Scientific Reports*, vol. 14, no. 1, p. 6312, 2024.

[57] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016.

[58] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

[59] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

[60] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, pp. 1398–1402 Vol.2, 2003.

[61] L. Zhang, H. Pan, B. Jia, L. Li, M. Pan, and L. Chen, "Lightweight dcgan and mobilenet based model for detecting x-ray welding defects under unbalanced samples," *Scientific Reports*, vol. 15, no. 1, p. 6221, 2025.

[62] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[63] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," *Advances in neural information processing systems*, vol. 31, 2018.

[64] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[65] M. Ferguson, R. Ak, Y.-T. T. Lee, and K. H. Law, "Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning," *Smart and sustainable manufacturing systems*, vol. 2, no. 1, pp. 137–164, 2018.

[66] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13*, pp. 740–755, Springer, 2014.

[67] H. Yu, X. Li, H. Xie, X. Li, and C. Hou, "Improving the industrial defect recognition in radiographic testing by pre-training on medical radiographs," *NDT E International*, vol. 149, p. 103260, 2025.

[68] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15750–15758, 2021.

[69] Z. Xie, Z. Zhang, Y. Cao, Y. Lin, J. Bao, Z. Yao, Q. Dai, and H. Hu, "Simmim: A simple framework for masked image modeling," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9653–9663, 2022.

[70] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*, pp. 1597–1607, PmLR, 2020.

[71] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9729–9738, 2020.

[72] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

[73] B. Trabucco, K. Doherty, M. Gurinas, and R. Salakhutdinov, "Effective data augmentation with diffusion models," *arXiv preprint arXiv:2302.07944*, 2023.

[74] H. Fang, B. Han, S. Zhang, S. Zhou, C. Hu, and W.-M. Ye, "Data augmentation for object detection via controllable diffusion models," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1246–1255, 2024.

[75] S. Lin, K. Wang, X. Zeng, and R. Zhao, "Explore the power of synthetic data on few-shot object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 638–647, 2023.

[76] Y. Tian, L. Fan, P. Isola, H. Chang, and D. Krishnan, "Stablerep: Synthetic images from text-to-image models make strong visual representation learners," in *Advances in Neural Information Processing Systems* (A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, eds.), vol. 36, pp. 48382–48402, Curran Associates, Inc., 2023.

[77] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*, pp. 8748–8763, PmLR, 2021.

[78] Y. Tian, L. Fan, K. Chen, D. Katabi, D. Krishnan, and P. Isola, "Learning vision from models rivals learning vision from data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15887–15898, 2024.

[79] R. A. Bafghi, N. Harilal, C. Monteleoni, and M. Raissi, "Mixdiff: Mixing natural and synthetic images for robust self-supervised representations," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 7500–7511, 2025.

[80] X. Xu, Z. Wang, G. Zhang, K. Wang, and H. Shi, "Versatile diffusion: Text, images and variations all in one diffusion model," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7754–7765, 2023.

[81] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *International conference on machine learning*, pp. 12310–12320, PMLR, 2021.

[82] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. M. Ni, and H.-Y. Shum, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection," *arXiv preprint arXiv:2203.03605*, 2022.

[83] D. P. Kingma, M. Welling, *et al.*, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.

[84] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[85] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.

[86] R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. H. Bermano, G. Chechik, and D. Cohen-Or, "An image is worth one word: Personalizing text-to-image generation using textual inversion," *arXiv preprint arXiv:2208.01618*, 2022.

[87] CVAT.ai Corporation, "Computer vision annotation tool (cvat)."

[88] SAE International, "Casting, classification of," 2010. AMS 2175.

[89] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[90] F. C. Akyon, S. Onur Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 966–970, 2022.

[91] L. Khemani, R. Yadav, and T. Doshi, "Novel optimized image enhancement and deep learning based crack detection algorithm in non-destructive testing," *Indian Journal of Science and Technology*, vol. 18, no. 7, pp. 487–503, 2025.

[92] F. L. de la Rosa, L. Moreno-Salvador, J. L. Gómez-Sirvent, R. Morales, R. Sánchez-Reolid, and A. Fernández-Caballero, "Improved surface defect classification from a simple convolutional neural network by image preprocessing and data augmentation," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pp. 23–32, Springer, 2024.

[93] G.-h. Yun, S.-j. Oh, and S.-c. Shin, "Image preprocessing method in radiographic inspection for automatic detection of ship welding defects," *Applied Sciences*, vol. 12, no. 1, 2022.

[94] OpenCV Developers, "cv::clahe contrast limited adaptive histogram equalization." OpenCV Documentation, 2024. `https://docs.opencv.org/4.x/d6/db6/classcv_1_1CLAHE.html`.

[95] G. Jocher, A. Chaurasia, and other contributors, "YOLOv8: Cutting-Edge Object Detection and Segmentation." `https://github.com/ultralytics/ultralytics`, 2023. Accessed: 2025-08-05.

[96] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[97] Y.-A. Lai, X. Zhu, Y. Zhang, and M. Diab, "Diversity, density, and homogeneity: Quantitative characteristic metrics for text collections," *arXiv preprint arXiv:2003.08529*, 2020.

APPENDIX A

EXTRA MATERIAL

A.1 CLIP-based Retrieval Results

A.1.1 Text Query

A.1.1.1 Shrinkage Cavity

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of shrinkage cavity in metal casting'

* clip similarity: 0.3523, caption: "Root Canal Before"
* clip similarity: 0.3488, caption: "Final X-ray of crowns on implants"
* clip similarity: 0.3479, caption: "cartridge bearing iwth radial load"
* clip similarity: 0.3467, caption: "A healthy implant that has had time to heal"
* clip similarity: 0.3462, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3453, caption: "Dental X-Ray"
* clip similarity: 0.3445, caption: "Radiografia intra-oral do 1° e 2° molares inferiores."
* clip similarity: 0.3441, caption: "Total Periodontal Bone Loss"
* clip similarity: 0.3440, caption: "Fig. 2 : Scan showing an about 1-mm-high sinus floor"
* clip similarity: 0.3435, caption: "Figure 18-2 Radiograph of maxillary molar area"
* clip similarity: 0.3421, caption: "Wisdom Tooth Woes | by tj.blackwell"
* clip similarity: 0.3420, caption: "Dental Operative Microscope and Retreatment, Finding Previously Underseen MB2, Root Canal Treatment Pre-Therapy"
* clip similarity: 0.3417, caption: "Figure 3. Radiograph of dental implant in place."
* clip similarity: 0.3416, caption: "Pre-operative radiograph for formocresol pulpotomy (Group b)"
* clip similarity: 0.3402, caption: "Closeup of a dental x-ray photo"
* clip similarity: 0.3401, caption: "Implants placed in jaw."
* clip similarity: 0.3401, caption: "Minimally Invasive Dental Implant Placement"
* clip similarity: 0.3396, caption: "mehes_fogaszat_szajsebeszet_All_on_4_dental_implants_2"
* clip similarity: 0.3391, caption: "Atypical Canal Configuration, Type V, Root Canal Treatment Post Therapy 48744-2"
* clip similarity: 0.3391, caption: "Implants can be used to replace a single tooth or a mouthful."

* clip similarity: 0.3384, caption: "Colorized x-ray showing three molar teeth with dental fillings"
* clip similarity: 0.3384, caption: "Dental x-ray"
* clip similarity: 0.3379, caption: "pre op and 9 years follow up ceramic implant keramik implantat"
* clip similarity: 0.3377, caption: "Porcelain crowns xray"
* clip similarity: 0.3371, caption: "X-ray of root canal"
* clip similarity: 0.3370, caption: "teeth scan"
* clip similarity: 0.3368, caption: "Pre-op X-ray of internal resorption and distal decay. December 2013."
* clip similarity: 0.3368, caption: "digital design of full arch implant work"
* clip similarity: 0.3361, caption: "FIG.23-17-cont'd F,Radiograph of tooth after removal of stabilizing appliance. G, Appearance of tooth 1 year after trauma. H, Radiograph of tooth 1 year after trauma. No root canal treatment was performed because of immature nature of apex at time of displacement. (Courtesy Dr. Peter Spalding, University of Nebraska, llncoln.)"
* clip similarity: 0.3360, caption: "X-Ray Showing Enamel Growth"

A.1.1.2 Shrinkage Sponge

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of shrinkage sponge in metal casting'

* clip similarity: 0.3546, caption: "Harnessing Diatoms to Produce Advanced Sensors"
* clip similarity: 0.3470, caption: "Standard scalpel edge through microscope"
* clip similarity: 0.3430, caption: "Spinachitina fragilis
Nagli 106 borehole, 679.40 m, Juuru Stage (272-148)"
* clip similarity: 0.3402, caption: "Shaman Black Charcoal Soap - face, hands & body (113 gr)"
* clip similarity: 0.3393, caption: "Clinique City Block Purifying Charcoal Brush"
* clip similarity: 0.3361, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3360, caption: "Figure 18-2 Radiograph of maxillary molar area"
* clip similarity: 0.3358, caption: "Root Canal Before"
* clip similarity: 0.3357, caption: "SEM Micrograph of NanoSurface Topography"
* clip similarity: 0.3345, caption: "Bumper Donut Buffer Black"
* clip similarity: 0.3332, caption: "Astronauts plug a 2-milimeter hole found on ISS due to micrometeorite"
* clip similarity: 0.3317, caption: "A healthy implant that has had time to heal"
* clip similarity: 0.3315, caption: "Image © Paula McCartney. Black Ice #1 and #2. Photograms"
* clip similarity: 0.3309, caption: "Spectral-domain optical coherence tomography of a patient with retinoblastoma demonstrating an absence of exudation, a corrugated hyperreflective band representing the inner aspect of the tumor, disruption of the surrounding retinal architecture, and a shadowing effect beneath the inner edge of the tumor."
* clip similarity: 0.3306, caption: "Diffraction Image"
* clip similarity: 0.3303, caption: "Minimally Invasive Dental Implant Placement"
* clip similarity: 0.3298, caption: "The Uncle Sam rough diamond crystal before cutting"

* clip similarity: 0.3295, caption: "Fig. 2 : Scan showing an about 1-mm-high sinus floor"
* clip similarity: 0.3294, caption: "black sponge on white background: minimal suprematism art by (b)ananartista sbuff"
* clip similarity: 0.3293, caption: "4” Crater Lake, Oregon, USA, Sandstone 3d printed Photo of actual print, North is up"
* clip similarity: 0.3293, caption: "This enhanced image shows the inside of a rimless pit about 180 meters (591 feet) in diameter, northwest of the mountain Ascraeus Mons in the northern hemisphere of Mars."
* clip similarity: 0.3291, caption: "Fig. 10. Lateral view of a detached gastropore lid from figure 9 (SEM, magnification x 200)."
* clip similarity: 0.3290, caption: "These images taken by NASA’s Voyager 2 show changes in the clouds around Neptune’s Great Dark Spot (GDS) over a four and one-half-day period. From top to bottom the images show successive rotations of the planet an interval of about 18 hours."
* clip similarity: 0.3287, caption: "SEM Specimen Preparation"
* clip similarity: 0.3286, caption: "The scans revealed ’hidden’ images on many of the stones. Here, arrow heads are clearly visible on stone 4 of Stonehenge."
* clip similarity: 0.3286, caption: "Cyanotype print made on an old photographic enlarger directly from 120 negative film without using a conventional contact printer and digital processing"
* clip similarity: 0.3285, caption: "X Ray Photograph - Dna Discovery by King’s College London Archives"
* clip similarity: 0.3285, caption: "X-ray Image Photograph - Dna Discovery by King’s College London Archives"
* clip similarity: 0.3285, caption: "X-ray Photograph - Dna Discovery by King’s College London Archives"
* clip similarity: 0.3285, caption: "Helix Photograph - Dna Discovery by King’s College London Archives"

A.1.1.3 Elongated Porosity

Top 30 CLIP-retrieval hits for prompt: ’a radiograph of elongated porosity in metal casting’

* clip similarity: 0.3496, caption: "Spinachitina fragilis
Nagli 106 borehole, 679.40 m, Juuru Stage (272-148)"
* clip similarity: 0.3491, caption: "Security cameras at Leigh Galbraith’s Coffs Harbour home captured some incredible night time footage."
* clip similarity: 0.3434, caption: "Harnessing Diatoms to Produce Advanced Sensors"
* clip similarity: 0.3367, caption: "Spectral-domain optical coherence tomography of a patient with retinoblastoma demonstrating an absence of exudation, a corrugated hyperreflective band representing the inner aspect of the tumor, disruption of the surrounding retinal architecture, and a shadowing effect beneath the inner edge of the tumor."
* clip similarity: 0.3353, caption: "This enhanced image shows the inside of a rimless pit about 180 meters (591 feet) in diameter, northwest of the mountain Ascraeus Mons in the northern hemisphere of Mars."
* clip similarity: 0.3345, caption: "The interface between the semiconductor and the metal is

perfect and establish the new superconducting hybrid crystals, which could ultimately form the basic for future superconducting electronics. Credit: University of Copenhaga"
* clip similarity: 0.3341, caption: "white stretchy spider web halloween decoration horror diy cobweb scary party supplies prop for bar haund home horror g782 halloween scene decoration props"
* clip similarity: 0.3341, caption: "Coronal oblique MIP of contrast enhanced MRA demonstrating asymmetric narrowed lumen of right vertebral artery with irregular contour."
* clip similarity: 0.3336, caption: "Fig. 10. Lateral view of a detached gastropore lid from figure 9 (SEM, magnification x 200)."
* clip similarity: 0.3335, caption: "Fig. 3 : More posterior scan"
* clip similarity: 0.3328, caption: "X-ray beam induced secondary electrom image"
* clip similarity: 0.3324, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3323, caption: "Fig. 2 : Scan showing an about 1-mm-high sinus floor"
* clip similarity: 0.3297, caption: "Root Canal Before"
* clip similarity: 0.3297, caption: "Microscope image of a piece of an endangered wood species. The sample was analyzed using LADI-MS, where the entire sample of wood could be analyzed whole"
* clip similarity: 0.3289, caption: "SEM Specimen Preparation"
* clip similarity: 0.3287, caption: "c4d human bones parts - Ultimate Human Spine (Part of)... by scyrus"
* clip similarity: 0.3286, caption: "A picture that shows the shock waves around a T-38 Talon aircraft on December 13, 1993."
* clip similarity: 0.3283, caption: "https://upload.orthobullets.com/topic/8014/images/Case E - femur shaft - xray ap - Parsons_moved.png"
* clip similarity: 0.3282, caption: "Figure 18-2 Radiograph of maxillary molar area"
* clip similarity: 0.3281, caption: "Standard scalpel edge through microscope"
* clip similarity: 0.3279, caption: "Sonar image shows the MV Cemfjord wreckage. Image credit: MAIB"
* clip similarity: 0.3277, caption: "The presence (left) or absence (right) of aerial hyphae in Streptomyces may influence their atmospheric H₂ consumption"
* clip similarity: 0.3277, caption: "Exhibits of the RIEDEL Museum Kufstein"
* clip similarity: 0.3273, caption: "Comb Plate for LG Escalator DSA200168C/D"
* clip similarity: 0.3272, caption: "X-ray - Ilizarov apparatus - Lower leg fracture"
* clip similarity: 0.3269, caption: "Atomic layer deposition results in a conformal thickness of material. The layer evenly and uniformly coats any surface geometry. Image courtesy of Thomas Proslier."
* clip similarity: 0.3266, caption: "X-Rays Of Women Wearing Corsets"
* clip similarity: 0.3263, caption: "A healthy implant that has had time to heal"
* clip similarity: 0.3261, caption: "X Ray Photograph - Dna Discovery by King's College London Archives"

A.1.1.4 Round Porosity

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of round porosity in metal casting'

* clip similarity: 0.3388, caption: "Clinique City Block Purifying Charcoal Brush"
* clip similarity: 0.3387, caption: "Harnessing Diatoms to Produce Advanced Sensors"
* clip similarity: 0.3377, caption: "X Ray Photograph - Dna Discovery by King's College London Archives"
* clip similarity: 0.3377, caption: "X-ray Image Photograph - Dna Discovery by King's College London Archives"
* clip similarity: 0.3377, caption: "X-ray Photograph - Dna Discovery by King's College London Archives"
* clip similarity: 0.3377, caption: "Helix Photograph - Dna Discovery by King's College London Archives"
* clip similarity: 0.3363, caption: "This enhanced image shows the inside of a rimless pit about 180 meters (591 feet) in diameter, northwest of the mountain Ascraeus Mons in the northern hemisphere of Mars."
* clip similarity: 0.3360, caption: "Security cameras at Leigh Galbraith's Coffs Harbour home captured some incredible night time footage."
* clip similarity: 0.3358, caption: "SEM Micrograph of NanoSurface Topography"
* clip similarity: 0.3352, caption: "Spinachitina fragilis
Nagli 106 borehole, 679.40 m, Juuru Stage (272-148)"
* clip similarity: 0.3350, caption: "The first image NASA's Perseverance rover sent back after touching down on Mars on Feb. 18, 2021. The view, from one of Perseverance's Hazard Cameras, is partially obscured by a dust cover. (NASA / JPL-Caltech)"
* clip similarity: 0.3331, caption: "Astronauts plug a 2-milimeter hole found on ISS due to micrometeorite"
* clip similarity: 0.3324, caption: "Magnetic Resonance Imaging of a sharp disk, using a traditional method (left) and a new method based on Compressed Sensing and Bregman iteration (right), from work of T. Goldstein."
* clip similarity: 0.3322, caption: "Spectral-domain optical coherence tomography of a patient with retinoblastoma demonstrating an absence of exudation, a corrugated hyperreflective band representing the inner aspect of the tumor, disruption of the surrounding retinal architecture, and a shadowing effect beneath the inner edge of the tumor."
* clip similarity: 0.3320, caption: "Ruby Red Face Paint Stamps - Soccer Ball"
* clip similarity: 0.3314, caption: "Transcript of x-ray images of welds of pipelines to identify defective areas"
* clip similarity: 0.3312, caption: "3D printed scaffold for spinal repair"
* clip similarity: 0.3309, caption: "Agate Slice Natural Large 1pc"
* clip similarity: 0.3289, caption: "A mammogram image showing a tumor. Provided by Vanderbilt University Radiology Department."
* clip similarity: 0.3284, caption: "Bumper Donut Buffer Black"
* clip similarity: 0.3277, caption: "Fig. 10. Lateral view of a detached gastropore lid from figure 9 (SEM, magnification x 200)."
* clip similarity: 0.3268, caption: "Diatoms are microscopic algae that are sensitive to water quality, and their intricate glass cell walls are preserved in the lake sediment (mud). Their record in the sediment thus provides a record of water quality."

* clip similarity: 0.3264, caption: "European Comet Lander May Wake Up from Space Slumber"
* clip similarity: 0.3261, caption: "magnified atoms - looks like an uneven black and white and gray waffled surface"
* clip similarity: 0.3259, caption: "X-ray beam induced secondary electrom image"
* clip similarity: 0.3257, caption: "WindTech 1500 Series 1500-12 Small Size Foam Ball Windscreen 3/8in Black"
* clip similarity: 0.3252, caption: "ink stain: Ink Stain with the Symbol of Heart"
* clip similarity: 0.3248, caption: "Photoshop Ink Wash Brush 'Mush Agog'"
* clip similarity: 0.3242, caption: "Dust Visual - sixpaxk.fr"
* clip similarity: 0.3242, caption: ""Bedside Manner , 2007. Graphite on paper. 7"" x 6"".""

A.1.1.5 Inclusion

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of inclusion in metal casting'

* clip similarity: 0.3404, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3389, caption: "A healthy implant that has had time to heal"
* clip similarity: 0.3332, caption: "Wisdom Tooth Woes | by tj.blackwell"
* clip similarity: 0.3323, caption: "Image result for xrays"
* clip similarity: 0.3323, caption: "Coronal oblique MIP of contrast enhanced MRA demonstrating asymmetric narrowed lumen of right vertebral artery with irregular contour."
* clip similarity: 0.3320, caption: "Dental X-Ray"
* clip similarity: 0.3310, caption: "Maxillary second premolar, buccal aspect. Ten typical specimens are shown."
* clip similarity: 0.3302, caption: "A panoramic patient's view of jaws, teeth, joints and sinus's. The upper jaw has only two molars on each side and no wisdom teeth. The lower left (has an L at the bottom) has a vertical wisdom tooth. If you look closely, you can see a darker space behind the tooth. This is bone loss because of infection. This tooth needs to come out. On the other hand, the one on the right is laying down sideways. It is fully encircled by bone. This one stays but needs to be checked every few years because it can become cystic."
* clip similarity: 0.3280, caption: "Real x-ray of devitalization tooth stock photo"
* clip similarity: 0.3280, caption: "the loss of a distinct lamina dura and the granular texture of the bone pattern in this periapical film"
* clip similarity: 0.3280, caption: "Pre-operative radiograph for formocresol pulpotomy (Group b)"
* clip similarity: 0.3279, caption: "Harnessing Diatoms to Produce Advanced Sensors"
* clip similarity: 0.3275, caption: "Porcelain crowns xray"
* clip similarity: 0.3274, caption: ""4-1/2"" LAMB -3D CHOCOLATE CANDY MOLD""
* clip similarity: 0.3270, caption: "X-Ray Teeth Icon"
* clip similarity: 0.3268, caption: "Minimally Invasive Dental Implant Placement"
* clip similarity: 0.3263, caption: "X-rays reveal golf balls inside the carpet snake."
* clip similarity: 0.3262, caption: "x ray equipment: X-ray picture of human knee joint isolated on white background"
* clip similarity: 0.3262, caption: "ray tracing: X-ray picture of human knee joint isolated on

white background"
* clip similarity: 0.3261, caption: "X-ray scan of humans teeth Royalty Free Stock Images"
* clip similarity: 0.3261, caption: "Root Canal Before"
* clip similarity: 0.3259, caption: "Closeup of a dental x-ray photo"
* clip similarity: 0.3258, caption: "A panoramic fi lm of a patient with osteopetrosis; note the increased density of the"
* clip similarity: 0.3254, caption: "Atypical Canal Configuration, Type V, Root Canal Treatment Post Therapy 48744-2"
* clip similarity: 0.3249, caption: "X-Ray of humerus 3"
* clip similarity: 0.3248, caption: "Osteoid osteoma in proximal femur"
* clip similarity: 0.3248, caption: "Treatment of Class II division 1 malocclusion treatment with mandibular advancement features"
* clip similarity: 0.3247, caption: "Figure 3. Radiograph of dental implant in place."
* clip similarity: 0.3244, caption: "Dental Operative Microscope and Retreatment, Finding Previously Underseen MB2, Root Canal Treatment Pre-Therapy"
* clip similarity: 0.3243, caption: "Dental-Implants-Advanced-Technology-2D-Image"

A.1.1.6 Gas Hole

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of gas hole in metal casting'

* clip similarity: 0.3385, caption: "3D printed scaffold for spinal repair"
* clip similarity: 0.3379, caption: "Black aluminum oxide (black fused alumina) powder"
* clip similarity: 0.3360, caption: "Allstar Performance 56110 Nut, 1-8 in Thread, Steel, Zinc Oxide, Allstar Jack Bolts, Each"
* clip similarity: 0.3359, caption: "Spherical bead"
* clip similarity: 0.3346, caption: ""H6018X1 18-Tooth, 60 Standard Roller Chain Finished Bore Sprocket (3/4"" Pitch, 1"" Bore) - Froedge Machine & Supply Co., Inc.""
* clip similarity: 0.3313, caption: ""RAIN SPINNING BOWL 2008. 150 mm diameter. Brittannia Silver, Moonstone Blue enamel, white marble. Photo: Simon Armitt""
* clip similarity: 0.3301, caption: "Round Graphite Ingot Mold For 5Oz Silver Melting / Small"
* clip similarity: 0.3282, caption: "Wisdom Tooth Woes | by tj.blackwell"
* clip similarity: 0.3280, caption: "Urogram X_ray of the pelvis of a male patient aged 62, showing the dynamics of bladder control to test the health of the prostate gland image 2 of 2. ..."
* clip similarity: 0.3273, caption: "Pinhole Diaphragm for Wide-Angle Superspot"
* clip similarity: 0.3262, caption: "Clutch plates steel 1.5mm (T30M002010)"
* clip similarity: 0.3260, caption: "Hammered Pewter 18-Inch Prep/Bar Sink"
* clip similarity: 0.3253, caption: "SATIN NICKEL"
* clip similarity: 0.3248, caption: "Clinique City Block Purifying Charcoal Brush"
* clip similarity: 0.3247, caption: "Bumper Donut Buffer Black"
* clip similarity: 0.3245, caption: "A2432 Outlet Valve Rubber Diaphragm Seal"
* clip similarity: 0.3245, caption: "Ford Brake Friction Disc"

* clip similarity: 0.3245, caption: "Lava Console Table Stainless Steel"
* clip similarity: 0.3234, caption: "SEM Micrograph of NanoSurface Topography"
* clip similarity: 0.3224, caption: "Black Mini Brot Magnet"
* clip similarity: 0.3223, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3215, caption: "Hex Nut, Iron"
* clip similarity: 0.3213, caption: "teeth scan"
* clip similarity: 0.3212, caption: "Minimally Invasive Dental Implant Placement"
* clip similarity: 0.3211, caption: "Lower Bearing Guard"
* clip similarity: 0.3206, caption: "Plates for furnaces"
* clip similarity: 0.3203, caption: "High Domed Pewter Button, 175"
* clip similarity: 0.3203, caption: "175 L, High Domed Pewter Button"
* clip similarity: 0.3202, caption: "Pre-operative radiograph for formocresol pulpotomy (Group b)"
* clip similarity: 0.3198, caption: "Figure 3. Radiograph of dental implant in place."

A.1.1.7 Crack

Top 30 CLIP-retrieval hits for prompt: 'a radiograph of crack in metal casting'

* clip similarity: 0.3417, caption: "Security cameras at Leigh Galbraith's Coffs Harbour home captured some incredible night time footage."
* clip similarity: 0.3370, caption: "Patient (dwarfism) with a 10 cm tibial lengthening (external fixator) followed by a 16 cm femoral lenghtening using the lengthening nail."
* clip similarity: 0.3357, caption: "Image © Paula McCartney. Black Ice #1 and #2. Photograms"
* clip similarity: 0.3354, caption: "Root Canal Before"
* clip similarity: 0.3348, caption: "Wisdom Tooth Woes | by tj.blackwell"
* clip similarity: 0.3347, caption: "Microscope image of a piece of an endangered wood species. The sample was analyzed using LADI-MS, where the entire sample of wood could be analyzed whole"
* clip similarity: 0.3343, caption: "X-ray - Ilizarov apparatus - Lower leg fracture"
* clip similarity: 0.3325, caption: "Edge Break Windshield Crack"
* clip similarity: 0.3324, caption: "Nickel-base superalloy crack"
* clip similarity: 0.3316, caption: "Figure 18-2 Radiograph of maxillary molar area"
* clip similarity: 0.3302, caption: "Coronal oblique MIP of contrast enhanced MRA demonstrating asymmetric narrowed lumen of right vertebral artery with irregular contour."
* clip similarity: 0.3301, caption: "Fig. 2 : Scan showing an about 1-mm-high sinus floor"
* clip similarity: 0.3299, caption: "The presence (left) or absence (right) of aerial hyphae in Streptomyces may influence their atmospheric H₂ consumption"
* clip similarity: 0.3296, caption: "Dental X-Ray"
* clip similarity: 0.3290, caption: "Twin flame feathers and reflection classic round sticker"
* clip similarity: 0.3285, caption: "Spectral-domain optical coherence tomography of a patient with retinoblastoma demonstrating an absence of exudation, a corrugated hyperreflective band representing the inner aspect of the tumor, disruption of the surrounding retinal architecture, and a shadowing effect beneath the inner edge of the tumor."

* clip similarity: 0.3283, caption: "The Uncle Sam rough diamond crystal before cutting"
* clip similarity: 0.3282, caption: "Figure 3. Radiograph of dental implant in place."
* clip similarity: 0.3281, caption: "image for Goody's wedding dress 'turning into a new Turin Shroud'"
* clip similarity: 0.3280, caption: "femur with cancer"
* clip similarity: 0.3277, caption: "Texture Wall Cracked by Ivette-Stock"
* clip similarity: 0.3274, caption: "teeth scan"
* clip similarity: 0.3271, caption: "Cyanotype print made on an old photographic enlarger directly from 120 negative film without using a conventional contact printer and digital processing"
* clip similarity: 0.3267, caption: "A healthy implant that has had time to heal"
* clip similarity: 0.3266, caption: "Shaman Black Charcoal Soap - face, hands & body (113 gr)"
* clip similarity: 0.3262, caption: "white stretchy spider web halloween decoration horror diy cobweb scary party supplies prop for bar haund home horror g782 halloween scene decoration props"
* clip similarity: 0.3259, caption: "laser weld cross section; with wobble"
* clip similarity: 0.3259, caption: "B/W 1953 close up x-ray of bones in human neck + head moving up + down"
* clip similarity: 0.3251, caption: "The first image NASA's Perseverance rover sent back after touching down on Mars on Feb. 18, 2021. The view, from one of Perseverance's Hazard Cameras, is partially obscured by a dust cover. (NASA / JPL-Caltech)"
* clip similarity: 0.3250, caption: "Treatment of Class II division 1 malocclusion treatment with mandibular advancement features"

A.1.2 Image Query

A.1.2.1 Shrinkage Cavity

Top 30 hits for image query:

* clip similarity: 0.3462, caption: "Cloud of milk (cliccath) Tags: portrait bw cloud milk child noiretblanc dream lait nuage enfant lili rve canoneos5dmarkii canonef100mmf28macrousmlens cliccath cathschneider"
* clip similarity: 0.3382, caption: "Glaucous-winged Gull (Larus glaucescens) in flight, first winter plumage, Brickyards Beach, Gabriola Island, British Columbia, Canada"
* clip similarity: 0.3354, caption: "A sea otter mother floats alongside her days-old pup through the water. The pup still has the fluffy fur it was born with, which traps so much fur the pup cannot dive and floats like a cork, Enhydra lutris, Elkhorn Slough National Estuarine Research Reserve, Moss Landing, California"
* clip similarity: 0.3345, caption: "Fingerprint - detail of black fingerprint isolated on white"
* clip similarity: 0.3314, caption: "A bison emerges after swimming across the Yellowstone River in Yellowstone National Park, Wyoming, June 21, 2011. On average over 3,000 bison live in the park. Photo by: Jim Urquhart/Reuters (UNITED STATES)"
* clip similarity: 0.3308, caption: "Pigeons fly over the Liwa desert, 220 kms west of Abu Dhabi, on the sidelines of the Mazayin Dhafra Camel Festival on December 22, 2012. The

festival, which attracts participants from around the Gulf region, includes a camel beauty contest, a display of UAE handcrafts and other activities aimed at promoting the country's folklore. AFP PHOTO/KARIM SAHIB"
* clip similarity: 0.3300, caption: "Traces (Issa Fakhro) Tags: winter sky blackandwhite bw snow cold tree art beach nature beauty fog canon season landscape denmark photography skne europe flickr poetry photos sweden fineart footprints freezing philosophy chilly february scandinavia malm scania 2012 blackandwhitephotography artisticphotography resund sibbarp northerneurope 1635mm wideanglelense coldclimate resundsregionen followmeontwitterissafakhro issafakhro"
* clip similarity: 0.3297, caption: "Mute Swan (Cygnus Olor) floating on blue water in late winter at Emiquon National Wildlife Refuge near Lewistown in Fulton Co Illinois"
* clip similarity: 0.3294, caption: "Soprano Pipistrelle Bat. ©Laurie Campbell. NON SNH COPYRIGHT, FOR SNH USE ONLY ON THE WEBSITE. For information on reproduction rights contact the Scottish Natural Heritage Image Library on Tel. 01738 444177 or www.nature.scot"
* clip similarity: 0.3259, caption: "A duck hangs out in the secondary clarifier at the Springfield Metro Sanitary District's Spring Creek wastewater treatment plant, Friday, July 31, 2015, in Springfield, Ill. The waterfowl enjoy the water in the secondary clarifier due to how clean it has become at that point. Justin L. Fowler/The State Journal-Register"
* clip similarity: 0.3250, caption: "Wilson's storm-petrel, ~ 10 miles southeast of Chatham, Massachusetts, 6/24/12
 Canon 400mm f/5.6 on EOS 1D Mark IV
 1/2000 at f/8, ISO 400"
* clip similarity: 0.3244, caption: "FRED ZWICKY/JOURNAL STAR A cloud of frozen condensation rises off of a pond Monday, Feb. 3 near the Illinois River along Rt. 116 in East Peoria as a winter sun does little to warm ducks swimming in the water."
* clip similarity: 0.3244, caption: "White-faced storm petrel. Adult on ground showing back. Kundy Island, Stewart Island, March 2011. Image © Colin Miskelly by Colin Miskelly"
* clip similarity: 0.3241, caption: "Chihuahua on white background picture id696988706? b=1&k=6 &m=696988706&s=612x612&w=0&h=omkzb2r 1nxnt 8rkq1qihfkb0tgbfroi8pe0 goh7vu="
* clip similarity: 0.3236, caption: "Thousands of seagulls hover above the former Stinson Seafood Plant in Prospect Harbor late Thursday afternoon, February 18, 2010. Bumble Bee Foods, the current plant owner notified employees Wednesday, February 17 that the plant will close in April after being a fixture there for over 100 years. BANGOR DAILY NEWS FILE PHOTO BY JOHN CLARKE RUSS"
* clip similarity: 0.3221, caption: "© Licensed to London News Pictures. 29/01/2013. Westminster, UK A Cormorant catches and eats a fish in the River Thames next to Westminster Bridge this afternoon, 29th January 2013. Photo credit : Stephen Simpson/LNP"
* clip similarity: 0.3216, caption: "A townsend's big-eared bat (Corynothinus townsendii) exits a cave in the Derrick Cave complex, a series of lava tubes and lava bubbles. Dusk. Central Oregon. Please note: background elements have been digitally removed in this image."
* clip similarity: 0.3213, caption: "Eagle attack (vidular) Tags: nature nikon eagle wildlife baldeagle january iowa mississippiriver lightroom americanbaldeagle d90 haliacetusleucocephalus eagleattack lockanddam14 attackingeagle"
* clip similarity: 0.3212, caption: "Mountain hare footprints on the snow-covered slopes of Ben More Coigach, Wester Ross, December 2010 © Mark Hamblin/2020VISION"
* clip similarity: 0.3200, caption: "Chihuahua on white background picture id696988

692?b=1&k=6 &m=696988692&s=612x612&w=0&h=lpq8cur5e96arbhwidqqnjqntagm6pidy8d33xlihmq="
* clip similarity: 0.3199, caption: "Exposed (Issa Fakhro) Tags: winter sky blackandwhite bw snow cold tree art beach nature beauty fog canon season landscape denmark photography skne europe flickr poetry photos sweden fineart footprints freezing philosophy chilly february scandinavia malm scania 2012 blackandwhitephotography artisticphotography resund sibbarp northerneurope 1635mm wideanglelense coldclimate resundsregionen followmeontwitterissafakhro issafakhro"
* clip similarity: 0.3199, caption: "ID Smoke on the Water BW_9150"
* clip similarity: 0.3198, caption: "reconnaisance (lecates) Tags: fly d300 nikon seattle seagull iso200 10ev clouds 11600sec flight ferry gull 105mmf28gvrmicro pugetsound 105mm f8 sea sky"
* clip similarity: 0.3190, caption: "A stork flies over a flooded plain of the Oder river on May 24, 2010 in Lebus, eastern Germany, near the Polish German border. Public authorities expect the highest level of the flood wave to come to Germany on Wednesday or Thursday. The death toll from flooding in Poland rose to 15 Monday as torrential rain swelled major rivers to levels unseen in more than a century and rescuers from across Europe battled to prevent further tragedy."
* clip similarity: 0.3183, caption: "| Catharsis II forgotten dreams (Julia-Anna Gospodarou) Tags: trees winter blackandwhite bw white mountain snow nature monochrome square landscape highlights minimal greece negativespace zen dreamy serene highkey 2012 snowscape winterscape catharsis metsovo blackandwhitefineart nikond7000 juliaannagospodarou tamron18270pzd"
* clip similarity: 0.3183, caption: "Buzzard Landing in Snow Shower Photographic Print"
* clip similarity: 0.3181, caption: "Hovering (1963chris) Tags: lake bird nature birds sepia rural countryside flying wings movement raw wildlife gull sony blurred aves landing britishwildlife avian hovering slowshutterspeed blackheadedgull leightonmoss britishbirds mygear andme mygearandmepremium mygearandmebronze mygearandmesilver"
* clip similarity: 0.3172, caption: "The Heart Of A Rose (The Rosette Nebula in HST) (Terry Hancock www.downunderobservatory.com) Tags: camera sky monochrome night stars photography mono pier back backyard fotografie photos thomas space shed band science images astro apo m observatory telescope nebula astronomy imaging ccd universe rosette narrow cosmos constellation palette paramount hubble luminance the lodestar teleskop astronomie byo oiii refractor deepsky monoceros halpha ngc2237 ngc2244 astrograph autoguider starlightxpress ngc2238 ngc2239 Astrometrydotnet:status=solved ngc2246 Astrometrydotnet:version=14400 tmb92ss caldwell49 mks4000 qhy9m gt110s Astrometrydotnet:id=alpha20121141710324"
* clip similarity: 0.3170, caption: "Airborne (Lisa Franceski) Tags: ocean baby beach nature water sand nest wildlife flight chick airborne seabird shorebirds naturalhabitat commontern sternahirundo anawesomeshot commonternchick sigma120400mm canont2i commonternbaby lisafranceski photoofthedaynwf12"
* clip similarity: 0.3168, caption: "A duck spreads its wings to lift itself out of the ice that has formed around it on the Assiniboine Park Duck Pond Saturday morning. November 10 , 2012 (Ruth Bonneville/Winnipeg Free Press)"

A.1.2.2 Shrinkage Sponge

Top 30 hits for image query:

* clip similarity: 0.3541, caption: "Gray Cellular Shade Texture - Free High Resolution Photo"
* clip similarity: 0.3506, caption: "Gray Paper Texture - Free high resolution photo"
* clip similarity: 0.3503, caption: "Cotton 100% light gray cross, plus on a white background"
* clip similarity: 0.3489, caption: "smoke - colored smoke isolated on a white background"
* clip similarity: 0.3489, caption: "smoke - colored smoke isolated on a white background"
* clip similarity: 0.3465, caption: "blank old paper on a white background"
* clip similarity: 0.3462, caption: "Grey felt as background or texture Stock Photos"
* clip similarity: 0.3459, caption: "opaque: Set of opaque gray drops on white background"
* clip similarity: 0.3457, caption: "Squared sheet of paper, isolated on grey"
* clip similarity: 0.3455, caption: "Blank old paper on white background"
* clip similarity: 0.3443, caption: "Gray Canvas Texture, High Resolution"
* clip similarity: 0.3441, caption: "Gray Teflon Pan on White background Top view"
* clip similarity: 0.3438, caption: "Grunge graphite pencil texture on white background paper Stock Image"
* clip similarity: 0.3436, caption: "White Tube Isolated On A White Background"
* clip similarity: 0.3435, caption: "A rough texture background of white watercolour paper. High quality texture in extremely high resolution."
* clip similarity: 0.3434, caption: "Light gray paper texture or background"
* clip similarity: 0.3432, caption: "Infinite Rays of The Sun (1975-1978) Graphite on paper 19h x 23.25w in (48.26h x 59.06w cm) Collection of the Museum of Modern Art, New York"
* clip similarity: 0.3432, caption: "Smooth elegant white transparent cloth separated on grey background. Texture of flying fabric."
* clip similarity: 0.3427, caption: "A blank book cover on a gray background"
* clip similarity: 0.3425, caption: "blank poster on tripod on gray background"
* clip similarity: 0.3424, caption: "Blank grey newspaper on white background"
* clip similarity: 0.3422, caption: "Smooth elegant transparent white cloth separated on gray background. Texture of flying fabric."
* clip similarity: 0.3418, caption: "A coffee stain on a white background stock photo"
* clip similarity: 0.3415, caption: "Linen texture white for background (See my portfolio for more) - stock photo"
* clip similarity: 0.3413, caption: "white laboratory rat isolated on grey background."
* clip similarity: 0.3413, caption: "stain: coffee stain isolated on a white paper background"
* clip similarity: 0.3410, caption: ""Sunburned GSP #795(Clear skies descend into snow, North Slope, Alaska), 2014. Six 4""x5"" unique gelatin silver paper negatives. Private collection""
* clip similarity: 0.3406, caption: "A sheet of gray paper with a non-uniform color. Neutral paper texture."
* clip similarity: 0.3404, caption: "Close-up view of white rolled paper on white background"
* clip similarity: 0.3403, caption: "pattern of gray icicle on white background photo"

A.1.2.3 Elongated Porosity

Top 30 hits for image query:

* clip similarity: 0.3201, caption: "'More Rain' by M Ward by album review by Gregory Adams for The Full-length comes out March 4th via Merge Records."
* clip similarity: 0.3167, caption: "Daniel Avery SONG FOR ALPHA Vinyl Record"
* clip similarity: 0.3148, caption: "Recondite_album_artwork"
* clip similarity: 0.3134, caption: "Daniel Avery - Water Jump (Album Version) PHLP02"
* clip similarity: 0.3129, caption: "Clap Your Hands Say Yeah, Hysterical (CD)"
* clip similarity: 0.3125, caption: "ROTW: Daniel Avery - Love + Light"
* clip similarity: 0.3124, caption: "DANIEL AVERY – Song For Ewe"
* clip similarity: 0.3121, caption: "Recondite: Placid Vinyl 2LP"
* clip similarity: 0.3115, caption: "Daniel Avery - Drone Logic (Album Preview)"
* clip similarity: 0.3112, caption: "Daniel Avery - Divided Love Tape (Side B)"
* clip similarity: 0.3107, caption: "Daniel Avery - New Energy (Beyond The Wizard's Sleeve Re-Animation) [PH36]"
* clip similarity: 0.3103, caption: "Dan Croll Releases Debut Album 'Sweet Disarray' On 10th March 2014"
* clip similarity: 0.3102, caption: ""Daniel Avery: Song For Alpha + Bonus 10"" Vinyl""
* clip similarity: 0.3099, caption: "Daniel Avery - [Phase] / Conforce / Powell Remixes"
* clip similarity: 0.3098, caption: "Daniel Avery/SONG FOR ALPHA RMXS PT1 12""""
* clip similarity: 0.3096, caption: "Recondite - Placid (2LP)"
* clip similarity: 0.3096, caption: "Daniel Avery - Drone Logic - (CD)"
* clip similarity: 0.3096, caption: "CLAP YOUR HANDS SAY YEAH RELEASES EP"
* clip similarity: 0.3095, caption: "Dirty Projectors announce About To Die EP, share video"
* clip similarity: 0.3095, caption: "NEW MUSIC: M Ward – Migration Of Souls"
* clip similarity: 0.3094, caption: "Daniel Avery - Free Floating PHLP02"
* clip similarity: 0.3086, caption: "Daniel Avery / - All I Need / These Nights Never End (Remixes)"
* clip similarity: 0.3086, caption: "M. Ward Returns with 'More Rain'"
* clip similarity: 0.3083, caption: "Fresh Music: Daniel Avery - Drone Logic - Because Music"
* clip similarity: 0.3082, caption: "M WARD 'POST-WAR' CD"
* clip similarity: 0.3078, caption: "The Moon & Antarctica (2 LP 10th Anniversary Edition) by Modest Mouse at werd.com"
* clip similarity: 0.3077, caption: "GODSPEED YOU BLACK EMPEROR - Slow Riot For New Zero Kanada LP"
* clip similarity: 0.3076, caption: "M Ward - Migration Stories LP (Black)"
* clip similarity: 0.3074, caption: "Daniel Avery 'Together in Static' [PHLP14]"
* clip similarity: 0.3073, caption: "Damien Rice - My Favourite Faded Fantasy CD cover / CD borító 2014."

A.1.2.4 Round Porosity

Top 30 hits for image query:

* clip similarity: 0.3292, caption: "iamamiwhoami: BLUE | Album Reviews | Pitchfork"
* clip similarity: 0.3254, caption: "Dan Croll Releases Debut Album 'Sweet Disarray' On 10th March 2014"
* clip similarity: 0.3223, caption: "Recondite_album_artwork"
* clip similarity: 0.3222, caption: "Giraffage – Needs [Mixtape]"
* clip similarity: 0.3204, caption: "'More Rain' by M Ward by album review by Gregory Adams for The Full-length comes out March 4th via Merge Records."
* clip similarity: 0.3199, caption: "ROTW: Daniel Avery - Love + Light"
* clip similarity: 0.3196, caption: "Jamie Lidell - why_ya_why (taken from self-titled album 'Jamie Lidell' out Feb 18/19)"
* clip similarity: 0.3183, caption: "Hamilton Leithauser's new album, Black Hours, comes out June 3."
* clip similarity: 0.3171, caption: "Fresh Music: Daniel Avery - Drone Logic - Because Music"
* clip similarity: 0.3170, caption: "Daniel Avery - Drone Logic (Album Preview)"
* clip similarity: 0.3157, caption: "DANIEL AVERY – Song For Ewe"
* clip similarity: 0.3156, caption: "Neue Single: Cassels – A Snowflake In Winter"
* clip similarity: 0.3155, caption: "Hamilton Leithauser's new album, Black Hours, comes out June 3."
* clip similarity: 0.3155, caption: "Nouvel Album: White Noise Noah Gundersen"
* clip similarity: 0.3151, caption: "Daniel Avery - Water Jump (Album Version) PHLP02"
* clip similarity: 0.3150, caption: "Daniel Avery / - All I Need / These Nights Never End (Remixes)"
* clip similarity: 0.3148, caption: "Robin Guthrie/Harold Budd: White Bird in a Blizzard [Original Score] [Digipak]"
* clip similarity: 0.3145, caption: "Daniel Avery - Drone Logic - (CD)"
* clip similarity: 0.3139, caption: "'Indigo' by Wild Nothing, album review by Leslie Chu for Northern Transmissions"
* clip similarity: 0.3137, caption: "Albert Hammond Jr. 'Momentary Masters' (album stream)"
* clip similarity: 0.3133, caption: "Maxïmo Park: Nature Always Wins [Album Review]"
* clip similarity: 0.3133, caption: "RECENZJA: Lykke Li - I Never Learn"
* clip similarity: 0.3128, caption: "M. Ward Returns with 'More Rain'"
* clip similarity: 0.3122, caption: ""Lykke Li's ""So Sad So Sexy"" Album Announced With Tracklist & Release Date""
* clip similarity: 0.3119, caption: "NEW MUSIC: M Ward – Migration Of Souls"
* clip similarity: 0.3116, caption: "Slaughter Beach, Dog – Birdie - LP/CD - PREORDER"
* clip similarity: 0.3114, caption: "Daniel Avery - New Energy (Beyond The Wizard's Sleeve Re-Animation) [PH36]"
* clip similarity: 0.3111, caption: "Drake Bell - Honest (EP) - Album Download, Itunes Cover, Official Cover, Album CD Cover Art, Tracklist"
* clip similarity: 0.3108, caption: "new pornographers slide - The New Pornographers - Whiteout Conditions (Album Review)"
* clip similarity: 0.3105, caption: "Daniel Avery SONG FOR ALPHA Vinyl Record"

A.1.2.5 Inclusion

Top 30 hits for image query:

* clip similarity: 0.3511, caption: "Overhead photo of iceberg"
* clip similarity: 0.3490, caption: "An irregularly shaped ice cube is released on a white background."
* clip similarity: 0.3490, caption: "Cigarette stub isolated over a white background"
* clip similarity: 0.3471, caption: "heap of cigarette on a white background"
* clip similarity: 0.3468, caption: "human tooth on a white background"
* clip similarity: 0.3456, caption: "nasa b-31 iceberg satellite view"
* clip similarity: 0.3449, caption: "Ice cube on white background. File contains path to cut."
* clip similarity: 0.3448, caption: "Cigarette butt with ash over white background"
* clip similarity: 0.3445, caption: "Cigarette butt with ash over white background photo"
* clip similarity: 0.3437, caption: "Single grain of anise on white background"
* clip similarity: 0.3424, caption: "A white tooth on a white background"
* clip similarity: 0.3421, caption: "piece of cheese on a white background"
* clip similarity: 0.3421, caption: "Ice cube isolated on a white background"
* clip similarity: 0.3421, caption: "White vector paper tear placed on white background"
* clip similarity: 0.3417, caption: "Floating iceberg, showing its size under water"
* clip similarity: 0.3412, caption: "Photo of used teabag over white background"
* clip similarity: 0.3410, caption: "Small Oblong Shell On A White Background"
* clip similarity: 0.3408, caption: "Lit cigarette isolated on a white background"
* clip similarity: 0.3404, caption: "White tooth isolated on a white background"
* clip similarity: 0.3403, caption: "Cigarette isolated on a white background. Top view with copy space"
* clip similarity: 0.3402, caption: "cheese - A piece of cheese on a white background"
* clip similarity: 0.3401, caption: "Potato chips levitate on a white background."
* clip similarity: 0.3400, caption: "Top view of small band-aid on a white background"
* clip similarity: 0.3400, caption: "Extruded styrofoam on a white background,Extruded styrofoam sheet is isolated on a white background"
* clip similarity: 0.3396, caption: "Tooth protection isolated on a white background"
* clip similarity: 0.3393, caption: "Isolated broken cookie on a white background"
* clip similarity: 0.3393, caption: "Sugar isolated on the white background"
* clip similarity: 0.3390, caption: "Cubes of ice on a white background. File contains the path to cu - 52445702"
* clip similarity: 0.3388, caption: "Satellite image of the PIG iceberg (Image: Nasa)"
* clip similarity: 0.3386, caption: "broken: Two white broken cigarette on a white background"

A.1.2.6 Gas Hole

Top 30 hits for image query:

* clip similarity: 0.3556, caption: "Western Blotting (WB) image for anti-MAD2L1 (MAD2 Mitotic Arrest Deficient-Like 1 (Yeast)) (C-Term) (ABIN441319)"

* clip similarity: 0.3521, caption: "Rabbit IgG (H+L) Secondary Antibody (A16172) in Western Blot"
* clip similarity: 0.3517, caption: "Image no. 1 for Human Cell Line I Blot (ABIN2455957)"
* clip similarity: 0.3493, caption: "Dientamoeba fragilis trophozoites in a smear of pig feces after Giemsa staining, Italy, 2010–2011. Scale bar = 10 μm."
* clip similarity: 0.3485, caption: "Human 4-1BB, Mouse IgG2a Fc Tag, low endotoxin on SDS-PAGE under reducing (R) condition. The gel was stained overnight with Coomassie Blue. The purity of the protein is greater than 95%."
* clip similarity: 0.3463, caption: "Positive WB detected in:Mouse heart tissue,Human placenta tissue;All lanes:DOK3 antibody at 3?g/ml;Secondary;Goat polyclonal to rabbit IgG at 1/50000 dilution;Predicted band size: 54,24,36,25 KDa;Observed band size: 54 KDa;"
* clip similarity: 0.3446, caption: "Rabbit IgG Fc Cross-Adsorbed Secondary Antibody (A16124) in Western Blot"
* clip similarity: 0.3445, caption: "Image provided by the Independent Validation Program badge 29807.\{} \{} Lane 1: MCF7 cell lysate Lane 2:CAPAN1 cell lysates probed with Rabbit Anti-BRCA2 Polyclonal Antibody, Unconjugated (bs-1210R) at 1:200 for 10 hours at 4\{}u02daC. Followed by conjugation to secondary antibody at 1:10000 for 60 min at 37\{}u02daC."
* clip similarity: 0.3438, caption: "Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody (A16113) in Western Blot"
* clip similarity: 0.3435, caption: "Western Blotting (WB) image for anti-ATG4B antibody (ATG4 Autophagy Related 4 Homolog B (S. Cerevisiae)) (AA 23-53) (ABIN1449637)"
* clip similarity: 0.3413, caption: "ZNF428 Rabbit anti-Human, Polyclonal, Novus Biologicals 100μL; Unlabeled:Life"
* clip similarity: 0.3405, caption: "Western blot - Rabbit F(ab')2 Anti-Mouse IgG H&L (Alexa Fluor® 750) (ab175762)"
* clip similarity: 0.3402, caption: "Western blot of Galectin-9 antibody at 2 ug/ml with mouse heart tissue. Secondary: Goat polyclonal to Rabbit IgG at 1:15000 dilution. Predicted band size: 40 KDa. Observed band size: 40 KDa. This image was taken for the unconjugated form of this product. Other forms have not been tested."
* clip similarity: 0.3399, caption: "B2M / Beta 2 Microglobulin Antibody - Dot Blot results of rabbit Anti-Beta 2 Microglobulin Biotin Conjugated. Antigen: Beta-2-Microglobulin. Blot loaded at 3 fold dilution: 1. 100ng, 2. 33.3ng, 3. 11.1ng, 4. 3.70ng, 5. 1.23ng. Blocking: MB-070 Buffer for 30 minutes at RT. Primary Antibody: Rabbit Anti-Beta-2-Microglobulin Biotin 10μg/mL for 1hr at RT. Secondary Antibody: Streptavidin-HRP at 1:40,000 for 30min at RT. Imaging System ChemiDoc, Filter used: Chemi."
* clip similarity: 0.3392, caption: "ZNF385B Rabbit anti-Human, Polyclonal, Novus Biologicals 100μL; Unlabeled:Life"
* clip similarity: 0.3387, caption: "Positive Western Blot detected in HepG2 whole cell lysate, A549 whole cell lysate. All lanes: CFB antibody at 2.8 μg/ml Secondary Goat polyclonal to rabbit IgG at 1/50000 dilution. Predicted band size: 86 KDa. Observed band size: 86 KDa"
* clip similarity: 0.3383, caption: "CD163 Mouse anti-Human, Clone: OTI2B12, liquid, TrueMAB 100 μL; Unconjugated"
* clip similarity: 0.3381, caption: "C57Bl/6 splenocytes were stained with 0.25 ug PE-Cy7 Mouse Anti-CD4 ."
* clip similarity: 0.3380, caption: "Rabbit Anti-GRP78 Antibody used in Western blot (WB) on Rat Tissue lysates (SPC-107)"

* clip similarity: 0.3379, caption: "CDH11 Antibody (OAAF02843) in Jurkat cells using Western Blot"
* clip similarity: 0.3376, caption: "Western blot of human Jurkat T cell line lysate (1% laurylmaltoside); non-reduced sample, immunostained by mAb TRIM-04 and goat anti-mouse IgG (H+L)-HRP conjugate."
* clip similarity: 0.3375, caption: "Western blot of S35D1 Antibody with human Ovary Tumor lysate. This image was taken for the unconjugated form of this product. Other forms have not been tested."
* clip similarity: 0.3370, caption: "Fig. 1. Hepatozoon felis gamont within a neutrophil in a cat blood smear. Courtesy of Prof. Gad Baneth, School of Veterinary Medicine, Hebrew University, Jerusalem, Israel."
* clip similarity: 0.3368, caption: "Western Blotting Image 1: Bag1 (3.10G3E2) Mouse mAb"
* clip similarity: 0.3360, caption: "Western Blotting Image 1: Akt (pan) (40D4) Mouse mAb (Biotinylated)"
* clip similarity: 0.3357, caption: "COL25A1 Antibody (OAAF02925) in Jurkat, A549 cells using Western Blot"
* clip similarity: 0.3355, caption: "C57Bl/6 splenocytes were stained with 0.25 ug FITC Anti-Mouse CD4 (35-0042) (solid line) or 0.25 ug FITC Rat IgG2a (dashed line)."
* clip similarity: 0.3353, caption: "Paramecium caudatum Ciliate Protozoa with some in conjugation. LM X44."
* clip similarity: 0.3353, caption: "Western Blotting Image 1: Survivin (71G4B7) Rabbit mAb"
* clip similarity: 0.3353, caption: "Western Blotting (WB) image for anti-RASA4 (RAS P21 Protein Activator 4) (N-Term) (ABIN4349400)"

A.1.2.7 Crack

Top 30 hits for image query:

* clip similarity: 0.3507, caption: "Strand Of Hair On A White Background"
* clip similarity: 0.3446, caption: "Millau Viaduct - Fog , 1994, computer drawing printed on photo paper, cardboard"
* clip similarity: 0.3419, caption: "Untitled Barbed Wire Study II (1970) Graphite on paper 18.13h x 24.13w in (46.1h x 61.3w cm)"
* clip similarity: 0.3417, caption: "Tent (10) , 2018, graphite on cream paper, 14 x 17 inches"
* clip similarity: 0.3406, caption: "Elevated view of cotton twig on white backdrop"
* clip similarity: 0.3398, caption: "earthworm on a white background"
* clip similarity: 0.3394, caption: "Natural human hair on a white background"
* clip similarity: 0.3374, caption: "Tent (8) , 2018, graphite on cream paper, 14 x 17 inches"
* clip similarity: 0.3371, caption: "white mosquito on a gray background to represent mosquito control in thibadoux"
* clip similarity: 0.3366, caption: "Illustration-sketch of a black spider drawn in black china dangling isolated on a white sheet background"
* clip similarity: 0.3362, caption: "Image of Pink paper-clip on a white sheet, black background"
* clip similarity: 0.3362, caption: "Untitled barbed wire study (1970) Graphite on paper

18.13h x 24.13w in (46.1h x 61.3w cm)"
* clip similarity: 0.3359, caption: "Curtain 6, 2015, graphite on paper"
* clip similarity: 0.3358, caption: "an cidhe by Eòghann MacColl, Drawing, graphite on paper"
* clip similarity: 0.3357, caption: "Close Up Of Acoustic Guitar Strings. Instrument Lying On A White Background"
* clip similarity: 0.3353, caption: "A fragment of the tire tread, isolated on a white background"
* clip similarity: 0.3350, caption: "crack in the wall on a white background"
* clip similarity: 0.3344, caption: "Infinite Rays of The Sun (1975-1978) Graphite on paper 19h x 23.25w in (48.26h x 59.06w cm) Collection of the Museum of Modern Art, New York"
* clip similarity: 0.3338, caption: "Cloudbursting, 2015, Drawing, Pencil on paper, 29 x 42cm, $1.100"
* clip similarity: 0.3338, caption: "black and white image of vapour trail in cloudy sky over farm field"
* clip similarity: 0.3337, caption: "1984-85
Graphite on paper
 26.5 x 19.1 cm
Collection of the Artist"
* clip similarity: 0.3336, caption: "Untitled, 13 September 1977, pencil on paper, 21 x 29.7 cm, 1977"
* clip similarity: 0.3333, caption: "Title: Faint Drawing One Materials: Graphite powder, pearl pins, and white string on drywall."
* clip similarity: 0.3331, caption: "Image depicts a black line drawing of a shower head and valve/handle on a white background"
* clip similarity: 0.3330, caption: "The image of a clarinet isolated under the white background photo"
* clip similarity: 0.3329, caption: "Note with a paper clip. Isolated on a white background with a shadow - stock photo"
* clip similarity: 0.3329, caption: "close up of a white ripped piece of paper on white background"
* clip similarity: 0.3329, caption: "close up of a white ripped piece of paper on white background"
* clip similarity: 0.3328, caption: "worm on a white background"
* clip similarity: 0.3328, caption: "Study for Sudden Flight, 2012, Graphite on Paper (Stonehenge 100% Cotton), 11 x 14 in., by David Jay Spyker"