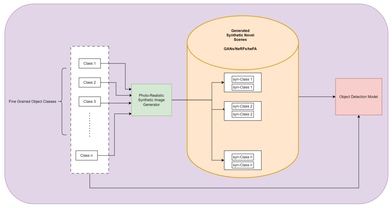
PHOTO-REALISTIC SYNTHETIC IMAGE GENERATION

STUDENTS: Fengrui Cheng, Jiahuan He, Nithin Philip Joseph, Fatwir SM, Tejoram Vivekanandan, Shikai Zhuang

Objective

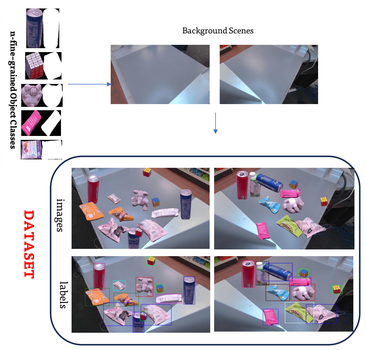
- A crucial challenge for computer vision applied to the retail industry is the **Automatic Checkout (ACO)** problem
- **Main challenges**: Seasonal, large-scale and fine-grained nature of the product categories as well as lack of large datasets
- We attempt to model a photorealistic synthetic image generation pipeline involving GANs, NeRFs and diffusion models to improve the Radius AI production classifier model, aiming for a higher accuracy



Synthetic Dataset

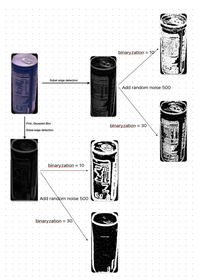
FEATURES:

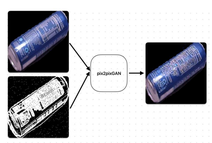
- **Seasonal Nature of Product and Data Annotation Efficiency:** Seasonal nature of products and cumbersome process of data annotation can make dataset curation a challenging process
- **Occlusion:** Simulate occlusions and overlapping items, relevant to automatic checkout scenarios
- **Realistic Environment:** Generation of retail environment helps train models to recognize and understand products in their actual context, including shelving, store layout, and visual cues
- **Baseline Dataset Features:**
  1. **2000/500** train/valid split
  2. **0.2 degree** of overlap between objects
  3. Maximum of **15 objects per scene**
  4. N**ormalized YOLO annotations/image**

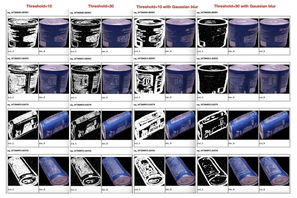


Synthetic Data Generation

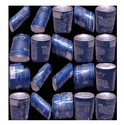
- **Pix2Pix GAN**: Enhances data variability and potentially improves accuracy
- **Inputs**: Utilize images and masks with the regular pix2pix architecture
- **Mask Generation**: Employ Sobel Edge Detector to generate masks, providing detailed information
- **Gaussian Blurring**: Remove noise and extract main features through Gaussian blurring





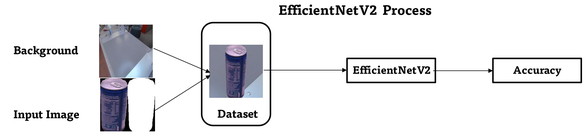


- **Modified DCGANs**: Learn underlying distribution of input data and generate realistic images
- **Generator**: Utilize U-Net architecture for better high-frequency features in generated images
- **Training Data**: Incorporate noise and mask for enhanced learning
- **Varying Image Scales**: Experimentation with (128,128) image scale yields sharper images



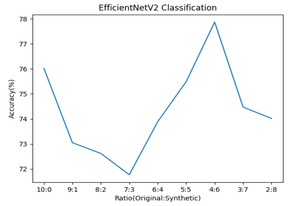
Object Level Classification [EfficientNetV2]

- **Training Data**: Utilize image masks, images, and the scene of the item image to generate synthetic input data
- **Validation Data**: Incorporate real item images as validation data
- **Model Training**: Employ EfficientNetv2 to classify different types of items and measure model accuracies
- **Image Padding Methods**: Explore various padding methods and select the one that yields the highest model accuracy
- **Data Ratio Adjustment**: Vary the ratio of original and generated data to test the model's classification accuracy
- **Result Analysis**: Assess the impact of data ratios on the model's performance



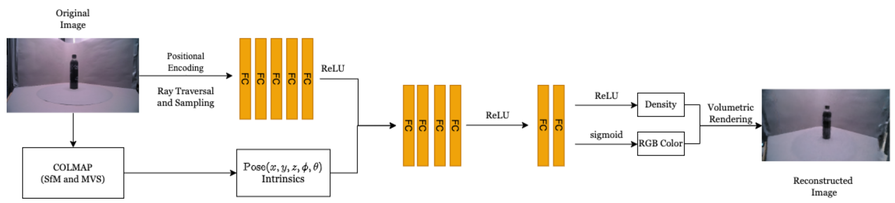
EfficientNetV2 Results

Proportion(original input : synthetic input)	original input	synthetic input	synthetic input(each class)	epochs	accuracy
10 : 0	2652	0	0	57	76.02%
9 : 1	2652	295	20	52	73.05%
8 : 2	2652	663	44	52	72.62%
7 : 3	2652	1137	76	45	71.77%
6 : 4	2652	1768	118	39	73.89%
5 : 5	2652	2652	177	37	75.49%
4 : 6	2652	3978	265	28	77.87%
3 : 7	2652	6188	413	12	74.47%
2 : 8	2652	10608	707	17	74.02%



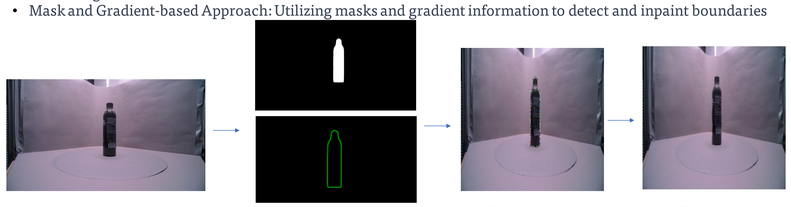
Neural Radiance Fields

- **Photorealistic Novel Views**: Transforming 2D images into 3D representations and then back to 2D
- **Efficient Scene Learning**: MLP (Multi-Layer Perceptron) overlearns scene details, optimizing storage and computation
- **Performance Evaluation**: Testing achieved a PSNR of **25.94 dB**, closely matching the original implementation



Diffusion Models

- Eliminating Synthetic Boundaries: Overcoming the presence of artificial boundaries between objects and background
- Mask and Gradient-based Approach: Utilizing masks and gradient information to detect and inpaint boundaries

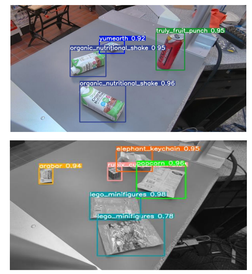


Results

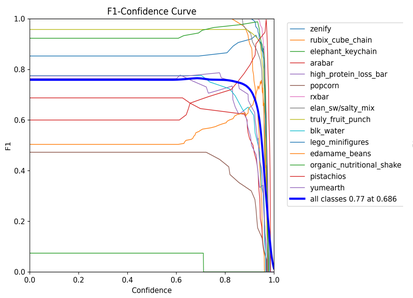
QUANTITATIVE RESULTS

Dataset -> Model	No of Images	mAP-50	mAP-50-95
Baseline Dataset/Model	1000	0.831	0.718
Baseline Dataset/Model	2000	0.859	0.768
syn-GAN-DATASETv1	2000	0.857	0.749
syn-GAN-DATASETv2	4000	0.882	0.765

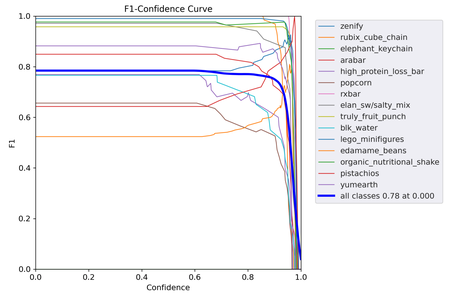
QUALITATIVE RESULTS



Baseline Results



Syn-GAN-DATASETv2 Results



Future Work, References, and Acknowledgments

- Explored GANs, NERFs, and Diffusion Models: Investigated various techniques, including pix2pix, DCGANs, and vanilla NERFs, for generating photorealistic synthetic images
- Discovered that object detection using greyscale images enhances the quality of deformable object detection
- Future Improvements: Future scope could be to enhance the quality of generated images by exploring advanced deep learning architectures and variants of GANs and NeRFs

- NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, Ben Mildenhall et. Al, ECCV 2020

- Acknowledgements: We express gratitude to TA Harsha, RadiusAI, and the ECE department for their invaluable insights, resources, and the opportunity to work on this remarkable project and gain valuable learning experiences

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

ADVISORS: DANIEL KING, ANDREA FANELLI, JENQ-NENG HWANG

SPONSORS: RADIUS AI