Seamless Human-Background Integration via Refined Instance Segmentation and Deep Image Blending

Ashka Shah, Sharique Pervaiz, Shivanshi Gupta
MScAC, University of Toronto

Motivation

The integration of moving human detection in indoor environments is crucial for enhancing virtual reality (VR) experiences by preventing unnecessary collisions and ensuring user safety. However, a significant challenge lies in the lack of publicly available indoor pedestrian datasets. To address this, our project focuses on developing a pipeline to transform outdoor datasets into indoor equivalents.

We identified a gap in existing image blending techniques, which rely heavily on hand-drawn masks—a time-consuming and subjective process. This inspired us to create an automated pipeline for data augmentation that eliminates the need for manual intervention. Our approach achieves seamless transformation from outdoor to indoor scenes, enabling the generation of robust datasets for indoor human detection.

By achieving the data augmentation phase, our project lays the groundwork for scalable and efficient dataset creation, paving the way for advancements in VR and indoor human detection technologies.

Related Work

Two gradient based seamless cloning algorithms Laplacian [1] & Poisson [2] blending have been widely in literature.

Laplacian Blending Technique

Step 1a. Build Laplacian pyramids LA and LB for images A and B respectively.
Step 1b. Build a Gaussian pyramid GR for the region image R.
Step 2. Form a combined pyramid LS from LA and LB using nodes of GR as weights. That is, for each l, i and j:

$$LS_l(i,j) = GR_l(i,j)LA_l(i,j) + (1 - GR_l(i,j))LB_l(i,j).$$

Step 3. Obtain the splined image S by expanding and summing the levels of LS.

Poisson Blending Technique

$$\min_f \iint_\Omega |\nabla f - \mathbf{v}|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \qquad (1)$$

Figure 3: Poisson equation put as an image interpolation problem

Poisson image blending works to smoothen the intensity of the blending boundary to make the transition seamless. It has a hard time adapting to the texture of the background image as it only focuses on the boundary.

Deep Image Blending

In 2020, Zhang et al introduce a new technique that blends a source image to a target image in two steps. The 1st step utilizes a differentiable loss that mimics the Poisson equation's objective along with content, style, TV (total variation) and histogram loss. The 2nd step optimizes the blended image on all the other losses except the Poisson based loss [3].

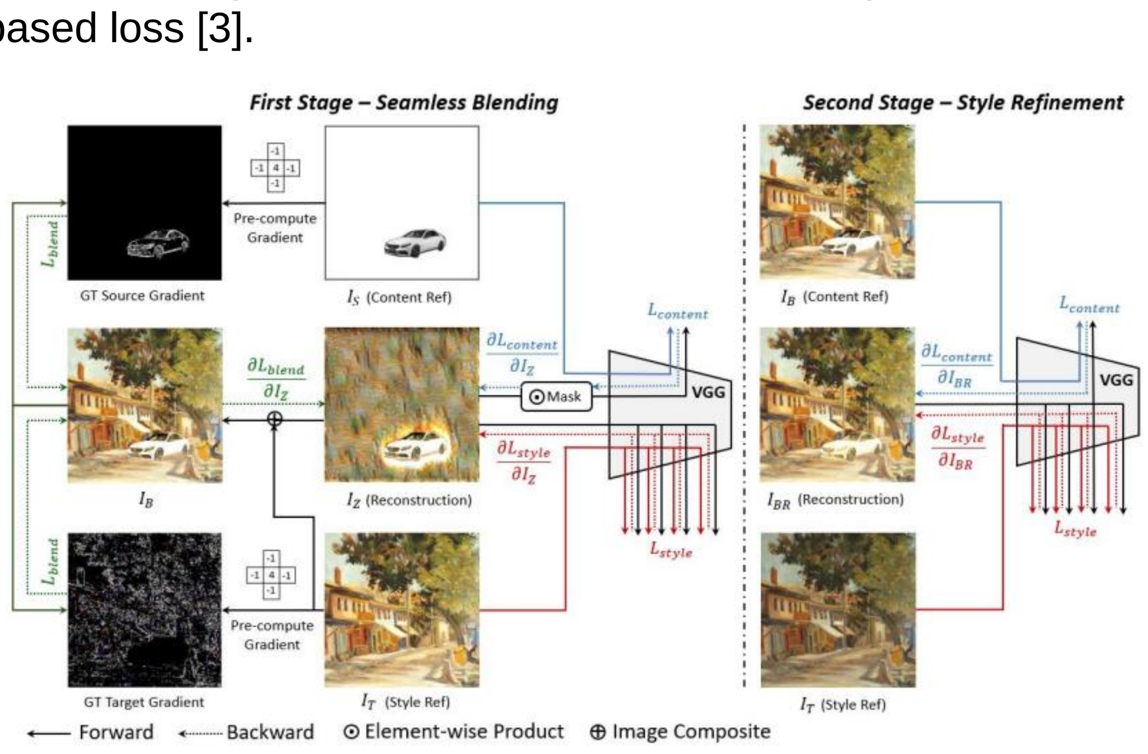


Figure 4: Deep Image Blending (DIB) Algorithm

References

[1] Burt, Peter J., and Edward H. Adelson. "A multiresolution spline with application to image mosaics." ACM Transactions on Graphics (TOG) 2.4 (1983): 217-236.
[2] Pérez, Patrick, Michel Gangnet, and Andrew Blake. "Poisson image editing." Seminal Graphics Papers: Pushing the Boundaries, Volume 2. 2023. 577-582.
[3] Zhang, Lingzhi, Tarmily Wen, and Jianbo Shi. "Deep image blending." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020.

New Technique

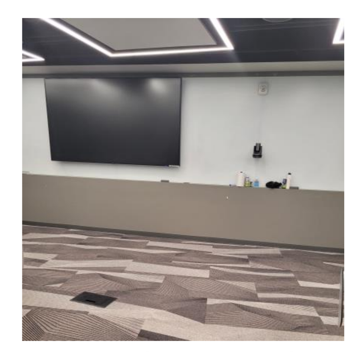
For integrating outdoor pedestrian images into indoor backgrounds with high realism and computational efficiency we propose combining mask refinement, telea inpainting and naïve light adjustment. This method can effectively replace the traditionally time-consuming and computationally intensive seamless-cloning step (Step 1) used in Deep Image Blending [3]. We combine mask refinement and telea inpainting with Step 2 of Deep Image Blending (DIB) to achieve this task.

Figure 1: Algorithm Design

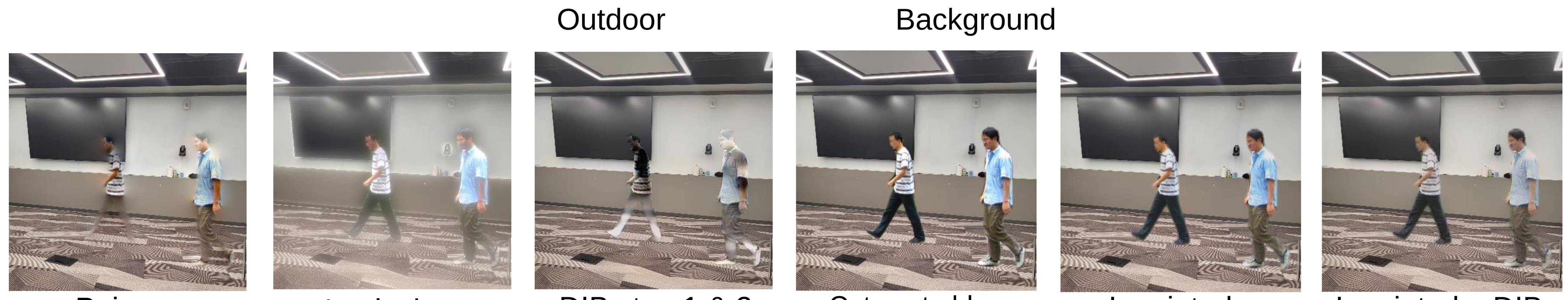
Experimental Results

Image Augmentation

Outdoor



Background



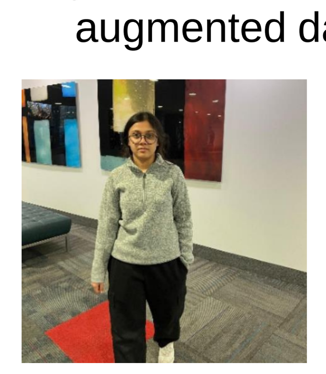
Poisson | Laplacian | DIB step 1 & 2 | Cut paste blur | Inpainted | Inpainted + DIB

Figure 5: Comparison of different blending techniques

	Models trained on different augmented datasets					
	Outdoors (baseline)	Poisson	Laplacian	Cutandpaste	Inpainted (Ours)	Inpainted+dib (Ours)
mAP IoU=0.50:0.95	0.478	0.534	0.581	0.592	0.627	**0.650**
mAP IoU=0.50	0.868	0.869	0.928	0.913	**0.937**	0.935
mAP IoU=0.75	0.523	0.621	0.670	0.763	**0.816**	0.793
mAR IoU=0.50:0.95	0.593	0.573	0.660	0.662	0.695	**0.702**

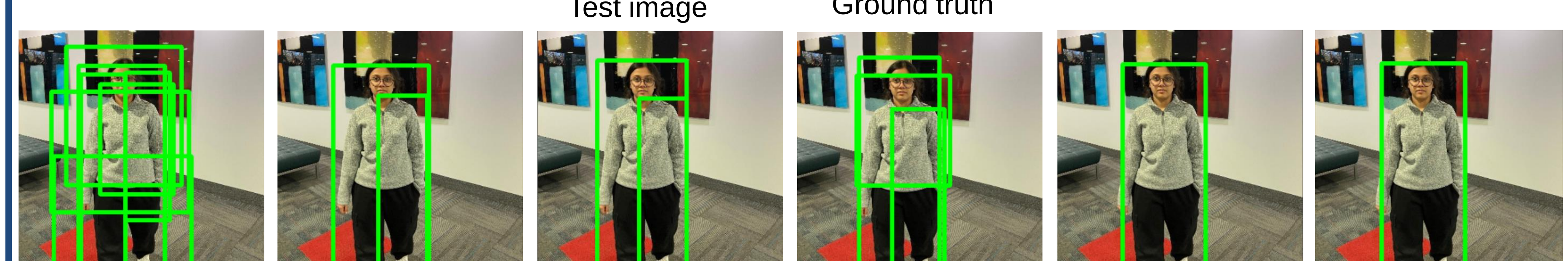
Table 1: Analysis of fasterRcnnResnet50 with different augmented data for training

Object Detection



Test image

Ground truth



Outdoors | Poisson | Laplacian | Cut paste blur | Inpainted (Ours) | Inpainted + DIB (Ours)

Figure 6: Bounding box comparisons of different models

VR Integration*

*Theoretical only. This is the intended use but not implemented due to pass through API not being public yet by Meta

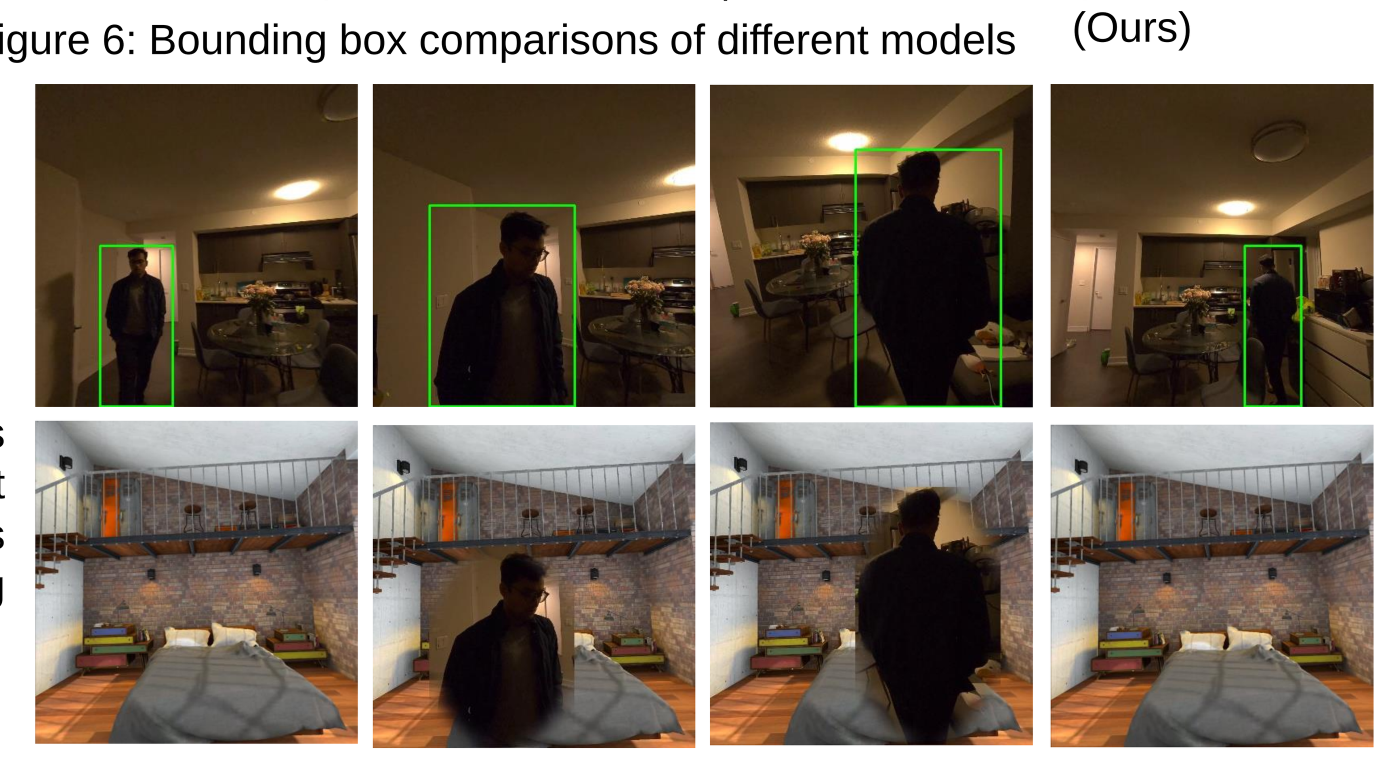


Figure 7: Example of showing passthrough when the person gets near the user