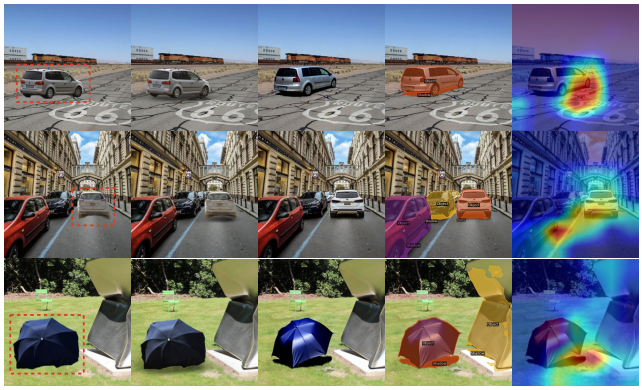
Shadows Don't Lie and Lines Can't Bend!
Generative Models don't know Projective Geometry...for now

Supplementary Material



Copy-Paste | Traditional Composites | ObjectStitch Composites | Object-Shadow (OS) | OS GradCam

Figure 8. Detecting Composite Errors with Object-Shadow (OS) Cues. We show images directly taken from Figure 1 (teaser) of [41] can be detected as generated using our object-shadow classifier. The bottom row provides a clear example where, despite the sun being positioned behind the camera, the shadows are mistakenly cast to the right. Also see the shadow of the adjacent object (marked in yellow), which is pointing upward in the opposite direction. Similarly, in the top row, shadows are cast in an implausible direction. The OS GradCam visualizations on the right successfully highlight these misdirected shadows.

9. Additional Analysis

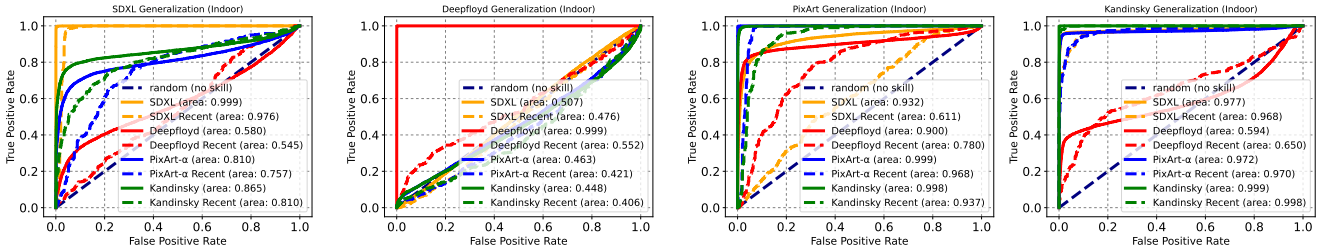
In Table 2, we provide quantitative analysis that our Line Segment cues and Perspective Field cues are correlated and look at similar geometric cues while Object-Shadow cues look for different geometric cues to identify if an image is generated or real.

We also provide statistical distributions of geometry cues leveraged for detecting projective geometry distortion. These include Object-Shadow pairs, Perspective Fields, and line segment distributions obtained from DeepLSD. The distributions are in Figures 17, 18, 19, 20, and 21.

An ROC plot in Figure 12 shows that while using statistical biases helps detect generated images over chance, ResNet classifiers trained directly on these cues still outperform them.

Table 2. We quantify the distribution of detection agreement among three types of cues: Line Segment (LS), Perspective Fields (PF), and Object-Shadow (OS), for the images processed by Stable Diffusion-XL. The output indicates whether each method can accurately identify generated images as either real or generated. The "Yes" indicates that the method has correctly detected generated images, whereas "No" indicates that the method has identified generated images as real. We have also provided the absolute and percentage values of images for both indoor and outdoor domains' unconfident test set in the last two columns. The table reveals a statistically significant correlation between Line Segment and Perspective field cues (p-value $\approx 2e^{-16}$), suggesting they are not independent in their detection of generated images. Conversely, Object-Shadow Cues demonstrate a different pattern of detection, with the probability of identifying an image as generated being lower than that of Line Segment Cues. This shows that they are complementary and look at distinct discrepancies in the images. A qualitative figure demonstrating a complementary capability is in Figure 6 of the main text.

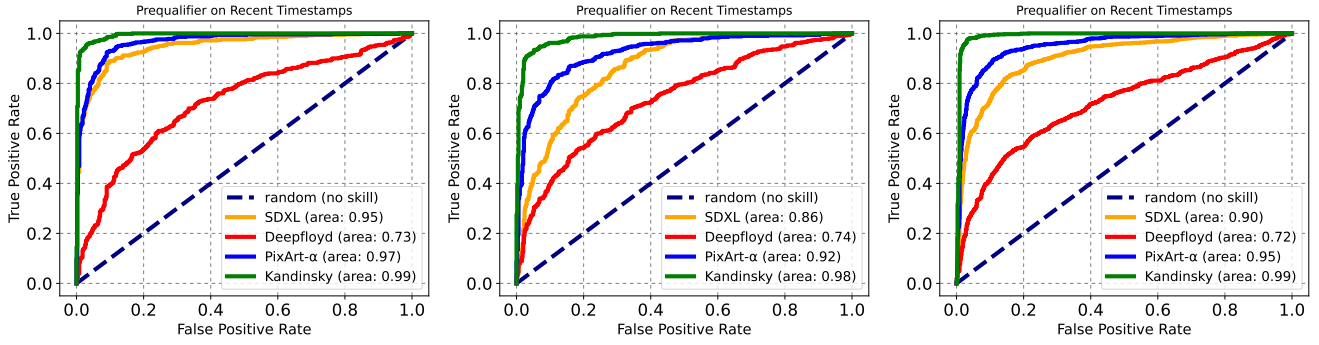
LS cues	PF cues	OS cues	Indoor	Outdoor
Yes	Yes	Yes	10520 (53.71%)	2382 (45.38%)
Yes	Yes	No	4844 (24.73%)	1314 (25.03%)
Yes	No	Yes	1033 (5.27%)	287 (5.47%)
Yes	No	No	725 (3.70%)	260 (4.95%)
No	Yes	Yes	872 (4.45%)	322 (6.13%)
No	Yes	No	874 (4.46%)	423 (8.06%)
No	No	Yes	285 (1.45%)	102 (1.94%)
No	No	No	435 (2.22%)	159 (3.03%)



(a) SDXL (Indoor) (b) Deepfloyd (Indoor) (c) PixArt-α (Indoor) (d) Kandinsky-v3 (Indoor)

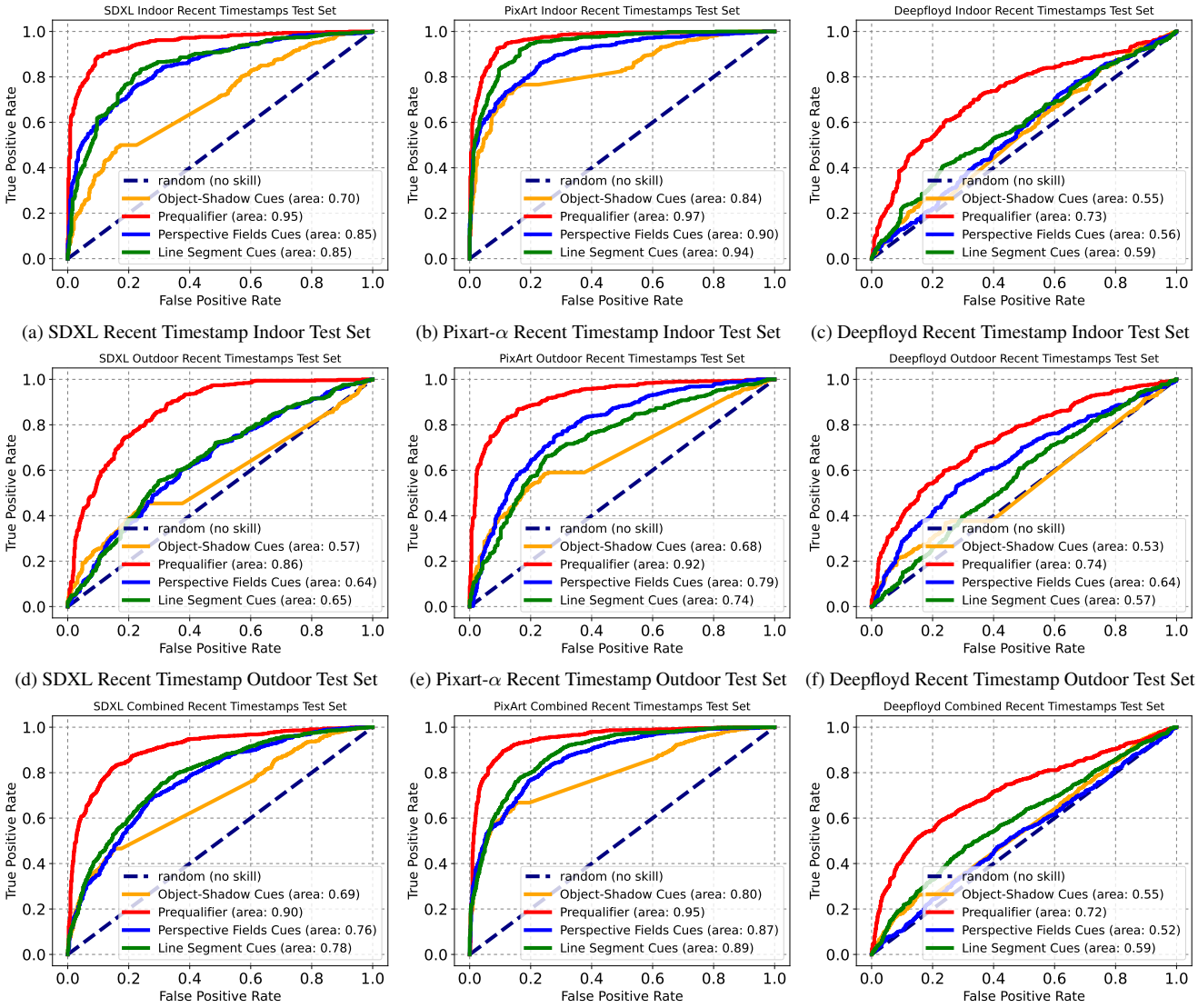
(e) SDXL (Outdoor) (f) Deepfloyd (Outdoor) (g) PixArt-α (Outdoor) (h) Kandinsky-v3 (Outdoor)

Figure 9. We trained classifiers on indoor and outdoor scenes using training sets consisting of images from different generators, following the indoor and outdoor splits depicted in Table 1. The models with the best validation accuracy over 30 epochs were selected. Data augmentation was performed using a protocol similar to that of [43], without blurring and with a JPEG compression probability of 5 percent, to improve generalizability towards both images generated from unseen generators and out-of-distribution real images with recent timestamps. The classifiers were evaluated on test sets containing 10,000 real images and 10,000 generated images from a target generator, paired caption-wise. Our results show that Kandinsky-v3 demonstrates the strongest generalization performance for both indoor and outdoor scenes. However, despite having the best validation accuracy, this Kandinsky-v3 classifier does not necessarily exhibit the absolute best generalization performance across all generators, possibly due to learning dataset-specific patterns. To address this, we selected a robust prequalifier that maintains comparable accuracy while generalizing more effectively to generators such as DeepFloyd. Also see Fig. 10.



(a) Indoor Prequalifer on Recent Timestamps (b) Outdoor Prequalifer on Recent Timestamps (c) Combined Prequalifer on Recent Timestamps

Figure 10. Based on the transferability experiments in Fig. 9, we chose to train our prequalifiers on Kandinsky-generated images. Instead of solely focusing on the highest validation accuracy, we selected prequalifiers that performed comparably well while demonstrating the best generalization towards other generators. This approach ensures the robustness of our prequalifiers. The resulting prequalifiers for indoor, outdoor, and combined settings all exhibit strong generalization performance on recent timestamp images from various generators, with particularly significant improved results on DeepFloyd-generated images.



(a) SDXL Recent Timestamp Indoor Test Set (b) Pixart-α Recent Timestamp Indoor Test Set (c) Deepfloyd Recent Timestamp Indoor Test Set

(d) SDXL Recent Timestamp Outdoor Test Set (e) Pixart-α Recent Timestamp Outdoor Test Set (f) Deepfloyd Recent Timestamp Outdoor Test Set

(g) SDXL Recent Timestamp Combined Test Set (h) Pixart-α Recent Timestamp Combined Test Set (i) Deepfloyd Recent Timestamp Combined Test Set

Figure 11. Our geometric classifiers, trained on derived geometric features from Kandinsky-v3 generated images, demonstrate strong generalization in detecting projective geometry errors within images generated by various unseen generators, using captions from real images with recent timestamps. We evaluate the classifiers on sets involving indoor scenes (top row), outdoor scenes (middle row), and a combination of indoor and outdoor scenes (last row). For indoor scenes, our perspective fields and line segment classifiers maintain strong AUCs greater than 0.84 for both SDXL and Pixart-α, while our object-shadow classifiers also exhibit comparable performance with AUCs of 0.70 and 0.84, respectively. Although the outdoor and combined settings pose a greater challenge compared to indoor scenes alone, our models, despite relying solely on geometric cues, remain robust towards SDXL and Pixart-α. However, generalization towards DeepFloyd proves to be quite challenging overall.

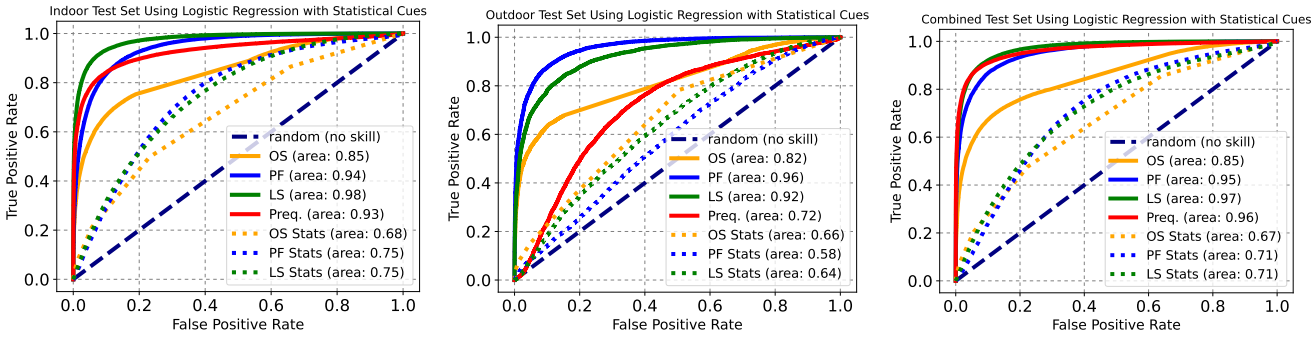
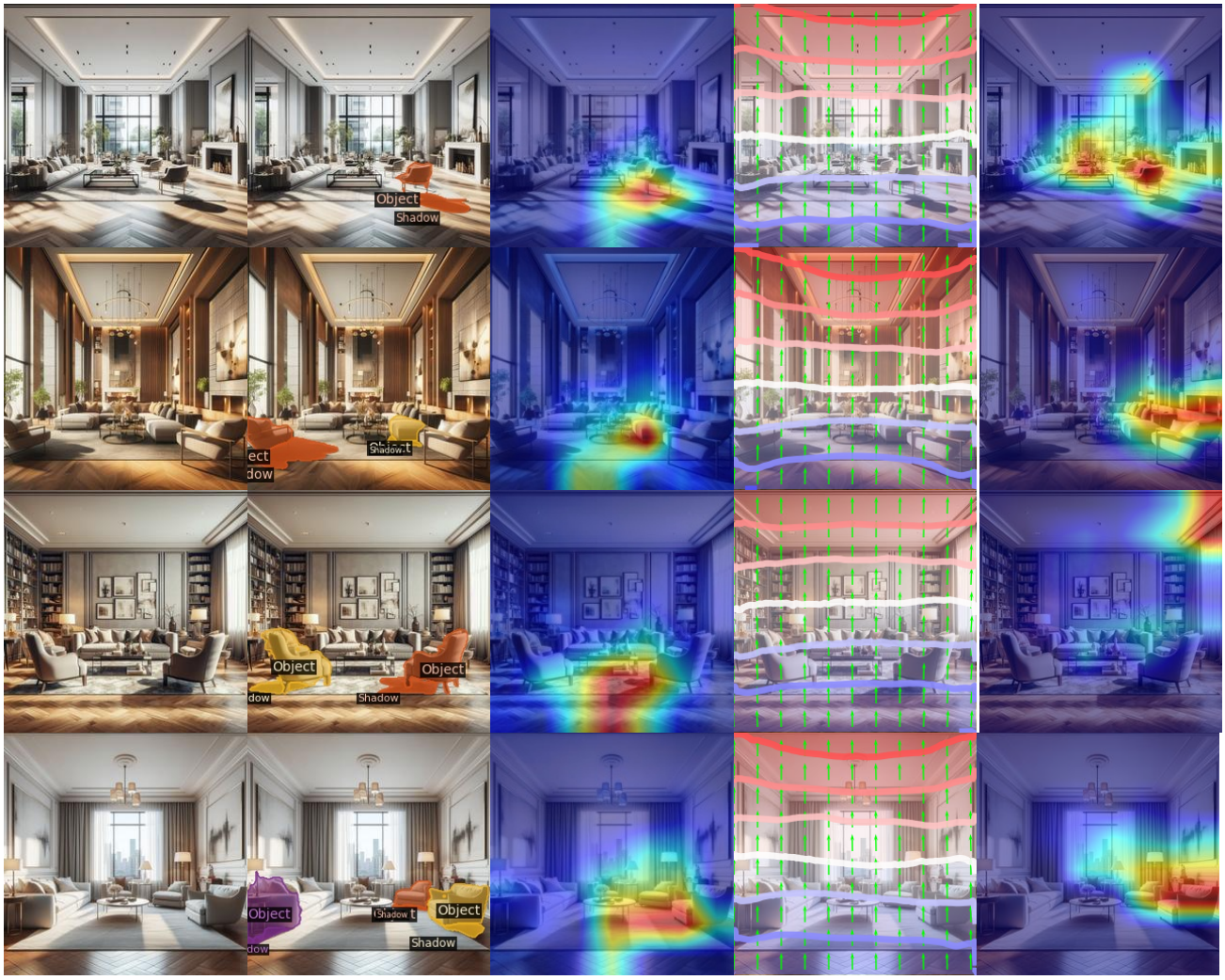


Figure 12. ROC analysis comparing our classifiers against basic statistical cues on our full test set. We compare the performance of our sophisticated classifiers – Object-Shadow (OS), Perspective Fields (PF) ResNet, and Line Segments (LS) PointNet classifiers – with basic statistical measures applied via logistic regression (LR) on indoor, outdoor, and combined test sets shown in dotted lines. While basic statistical cues like the count and mean lengths of line segments, the number of object shadows, and gravity changes per pixel indicate better-than-chance performance (AUCs ranging from 0.58 to 0.75), they are eclipsed by the more robust classifiers we developed and also the ResNet prequalifier. Our classifiers excel in identifying generated images with incorrect projective geometry by focusing on incorrect regions, not just statistical cues, as demonstrated by GradCAM visualizations.



Generated Image | Object-Shadow (OS) | OS GradCam | Perspective Fields | Perspective Fields GradCam

Figure 13. All interior scenes generated using Dalle-3. We analyze them using Object-Shadow (OS) cues and Perspective Fields (PF), along with their respective GradCam visualizations. The OS GradCam highlights areas where shadow directions or lengths don't appear to match the scene's lighting. For example, in the first and third rows, the shadows beneath the furniture don't seem to fit the objects casting them. The second row's OS GradCam shows an unnatural shadow on the sofa that's difficult to spot. Meanwhile, the PF analysis exposes inaccuracies in line alignment and vanishing points. In the top and third rows, the PF GradCam highlights inconsistencies along the room's ceiling lines and window frames that don't match the rest of the scene's perspective geometry. In the second and fourth rows, it detects inconsistencies on the side wall beneath the painting region.

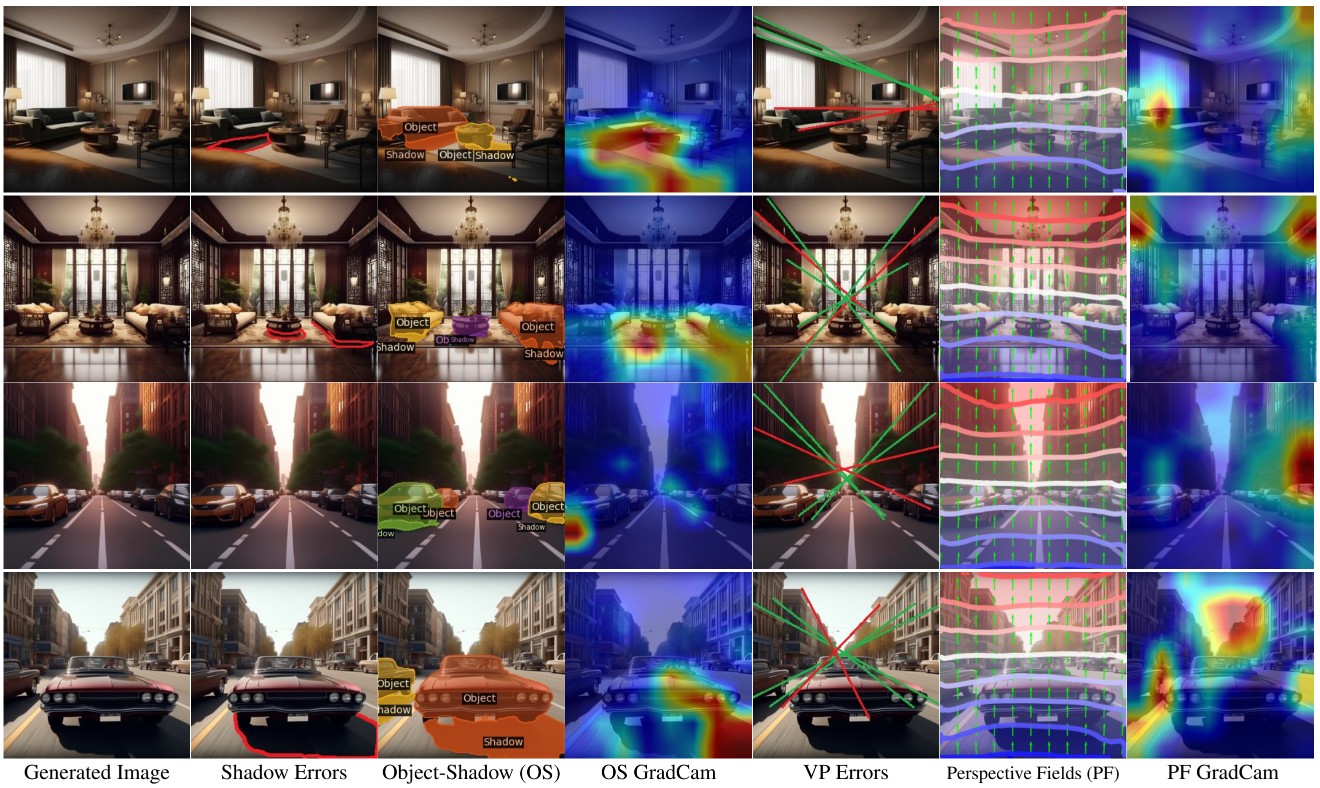
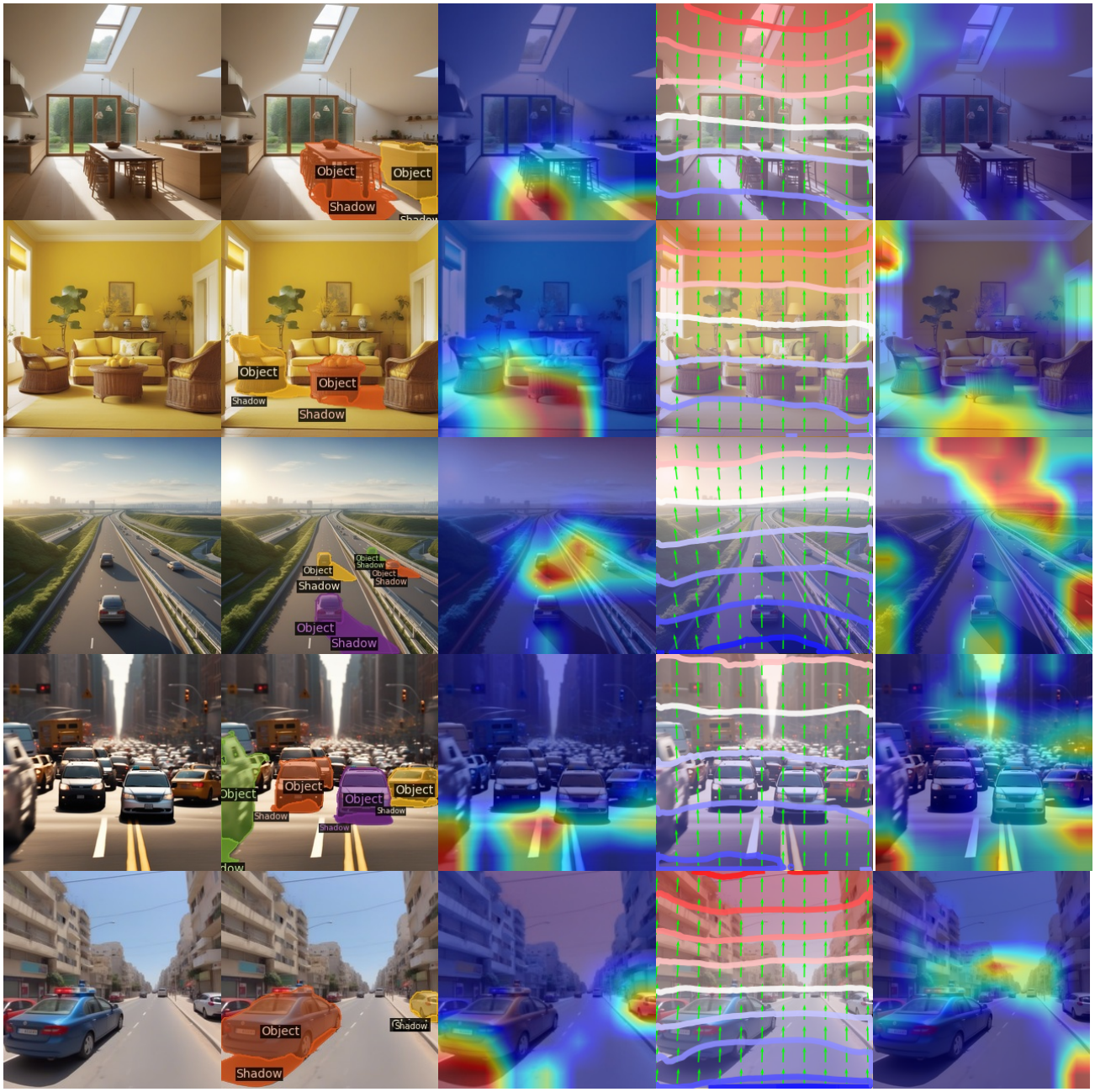


Figure 14. Grad-CAM results for indoor and outdoor scenes generated by Kandinsky. The second column shows shadow errors, while the third column overlays detected object-shadow pairs [44]. Grad-CAM applied to our Object-Shadow classifier (fourth column) reveals mismatched shadow lengths (first row), shorter-than-expected shadows (third row) and incorrect shadow shapes (fourth row). The sixth column shows Perspective Fields [21], and Grad-CAM applied to our Perspective Fields classifier (last column) confirms large perspective distortions on building facades, also supported by the vanishing point errors in the fifth column.



Generated Image | Object-Shadow (OS) | OS GradCam | Perspective Fields | Perspective Fields GradCam

Figure 15. The first column displays images generated by Stable Diffusion-XL. The second column overlays detected object-shadow pairs from [44], highlighting the model's ability to identify these features. The third column applies Grad-CAM to our Object-Shadow classifier. This shows areas most diagnostic of synthetic generation. Note: in the first row, the Grad-CAM weights suggest a shadow problem at the left side chair, which is difficult to check but plausible; in the second row, the shadow cast by the coffee table is in the wrong direction and Grad-CAM identifies this error as diagnostic. The fourth column shows the Perspective Fields of [21], and the fifth column shows Grad-CAM when applied to our Perspective Fields classifier. Note: in the first row, Grad-CAM weights identify a problem with the top of the cupboard on the left, which is difficult to confirm but plausible; in the second row, Grad-CAM weights identify a visible problem with the blind on the left. in the third row, the cars cast shadows in different directions and Grad-CAM identifies this error as diagnostic; in the fourth row, two cars in front cast shadows in different directions and Grad-CAM identifies this error as diagnostic; in the fifth row, Grad-CAM identifies the (very odd) structure of the buildings near the vanishing point as a problem, based on perspective field distortion. Best viewed on screen.

Generated Image | Object-Shadow (OS) | OS GradCam | Perspective Fields | Perspective Fields GradCam

Figure 16. The generated street scenes in Adobe's Firefly have inconsistencies in projective geometry. We show Object-Shadow (OS) and Perspective Fields (PF) analyses and have presented each generated image alongside the results. In the first row, the shadow of the bus on the left is in one direction, while the shadow of the bus on the right and the pedestrian point is in opposite directions. The second row shows the OS GradCam pinpointing a car's shadow that is unrealistically elongated on one side. In the third, we observe pedestrians with shadows that are inconsistent with the lighting. The Perspective Fields analysis in rows two and four detects line inconsistencies deep in the scene and near vanishing points, while in the first and last rows, it captures discrepancies on the road markings and building facades.

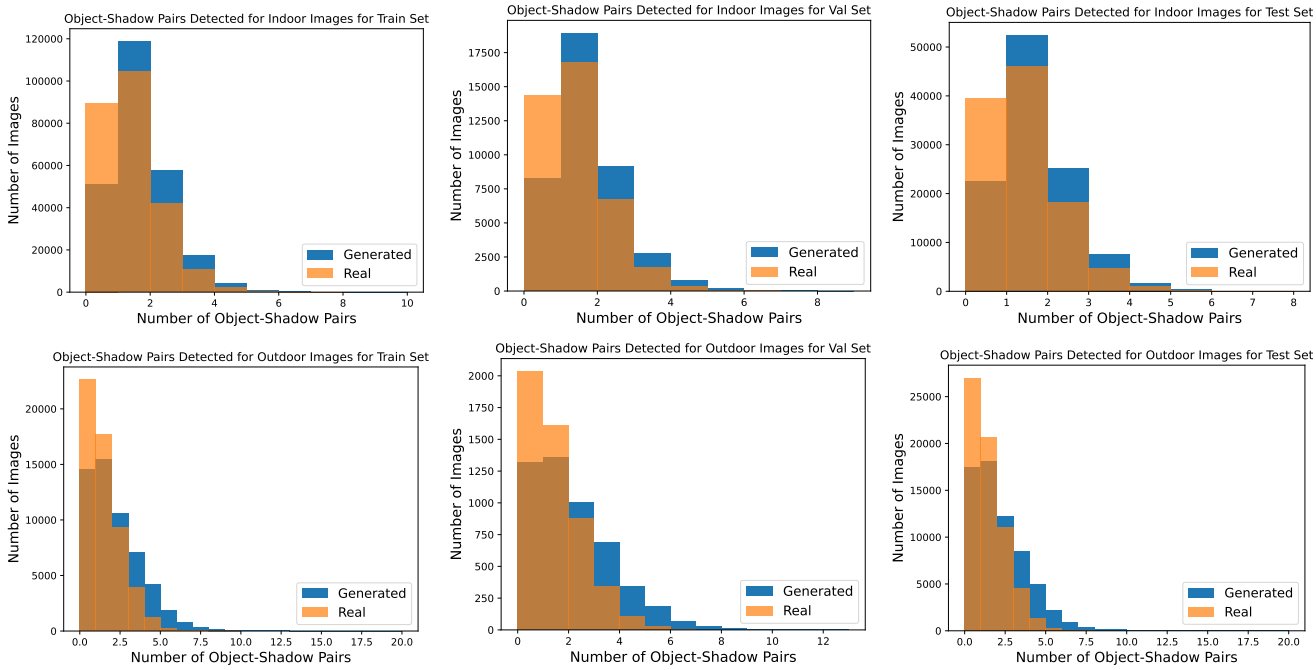


Figure 17. A statistical distribution analysis of a number of object-shadow pairs for both indoor and outdoor datasets. A classifier could exploit some of the statistical biases to distinguish between generated and real images. However, we found that our derived geometry cues perform much better than a classifier trained to look at such statistical signals, as shown in Figure 12. Furthermore, the GradCam analysis indicates that these derived object-shadow cues correctly identify erroneous regions.

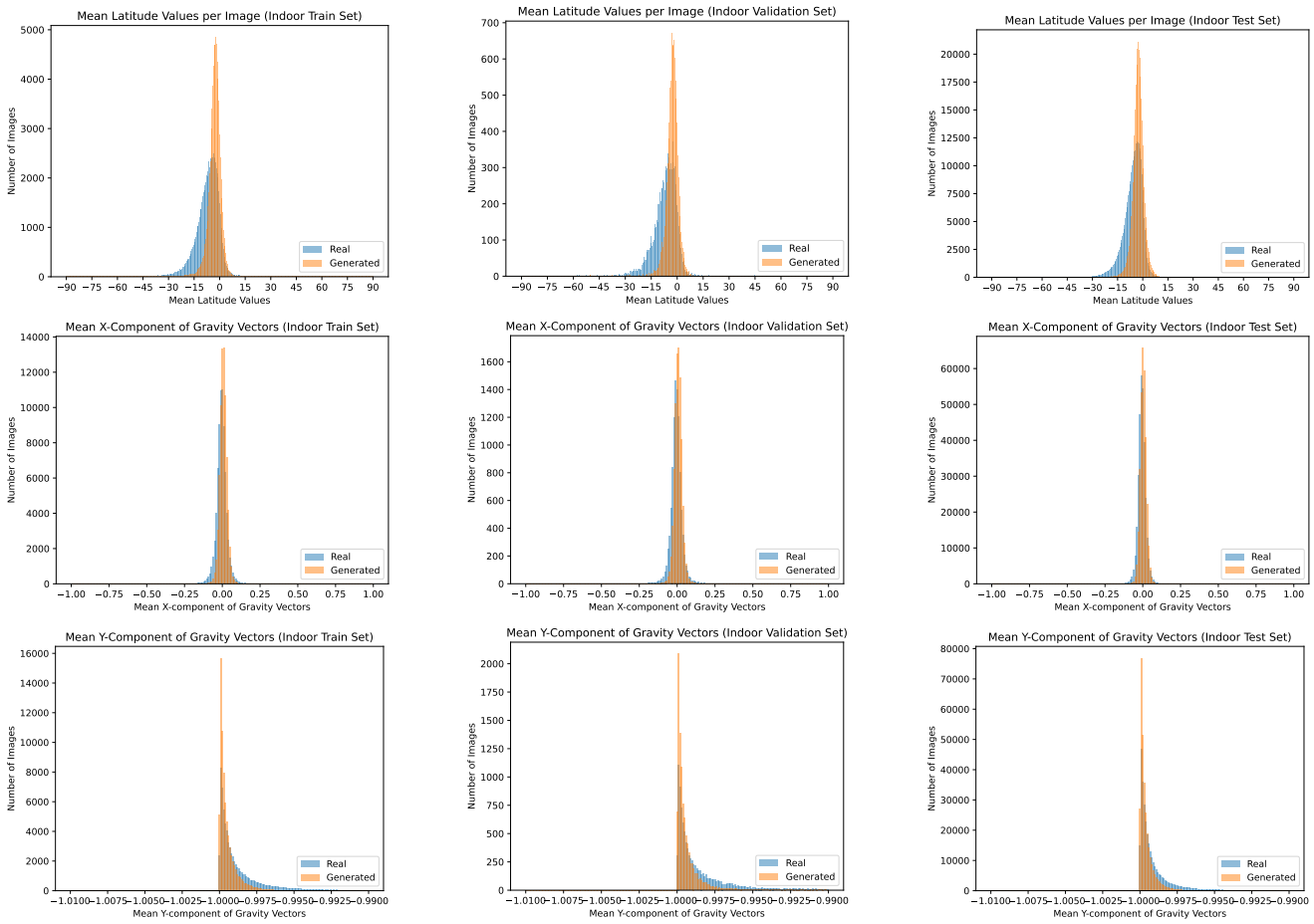


Figure 18. This set of histogram plots displays the statistical distribution of perspective field metrics in indoor scenes, comparing the training, validation, and test sets. The top row histograms reveal a significant difference in the distribution of latitude angles between real and generated images. The middle and bottom row plots illustrate the mean X and Y components of gravity vectors in the images, showing a clear separation between the real and generated images. These metrics indicate minor spatial inconsistencies between the real and generated images. Although these basic statistical differences provide some discriminative power, they are less effective than our ResNet classifier trained on Perspective Fields, which efficiently detects and focuses on critical geometric inconsistencies. This is validated by our comprehensive ROC analysis in Figure 12.

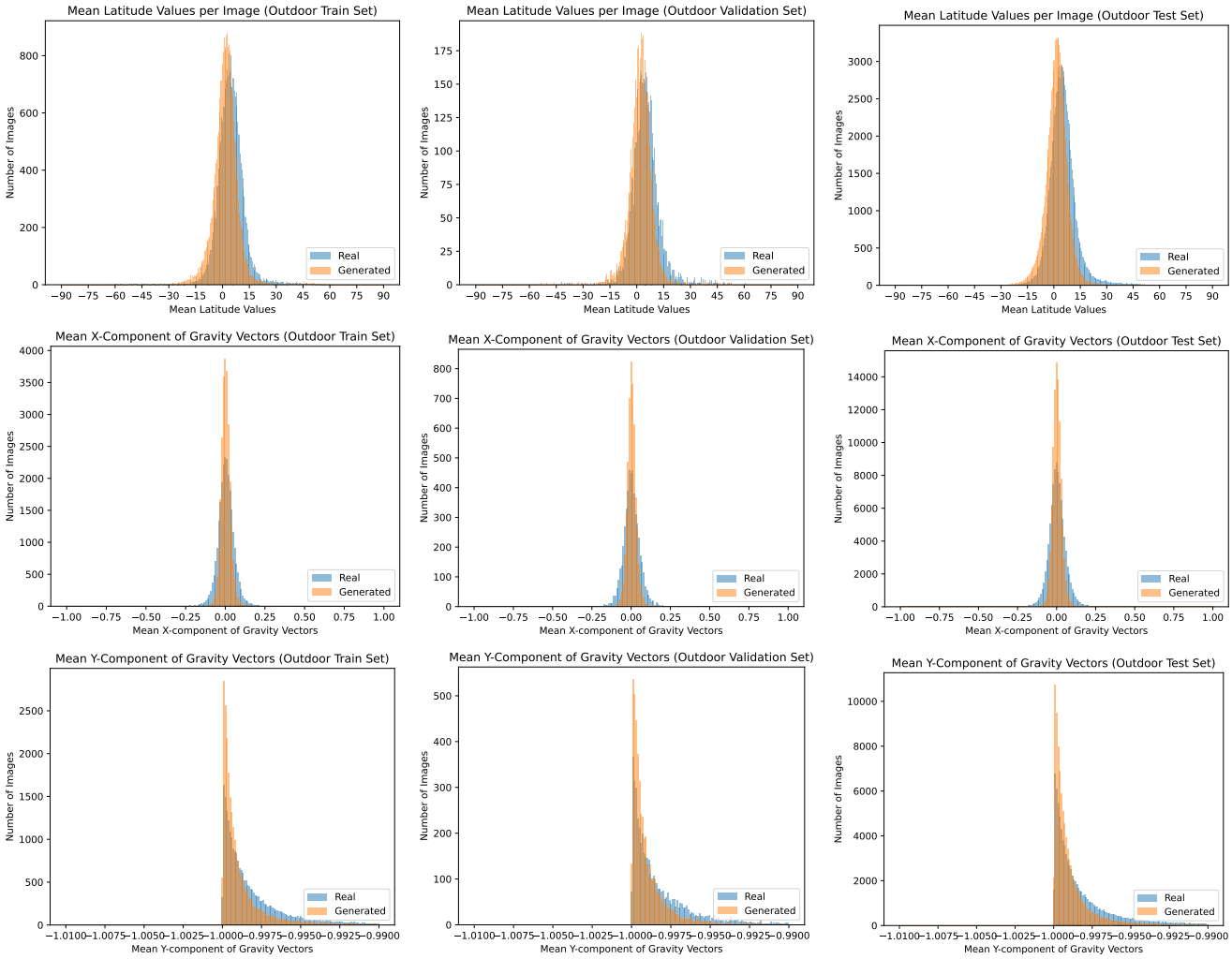


Figure 19. This set of histogram plots displays the statistical distribution of perspective field metrics in outdoor scenes, comparing the training, validation, and test sets. The top row histograms reveal a significant difference in the distribution of latitude angles between real and generated images. The middle and bottom row plots illustrate the mean X and Y components of gravity vectors in the images, showing a clear separation between the real and generated images. These metrics indicate minor spatial inconsistencies between the real and generated images. Although these basic statistical differences provide some discriminative power, they are less effective than our ResNet classifier trained on Perspective Fields, which efficiently detects and focuses on critical geometric inconsistencies. This is validated by our comprehensive ROC analysis in Figure 12.

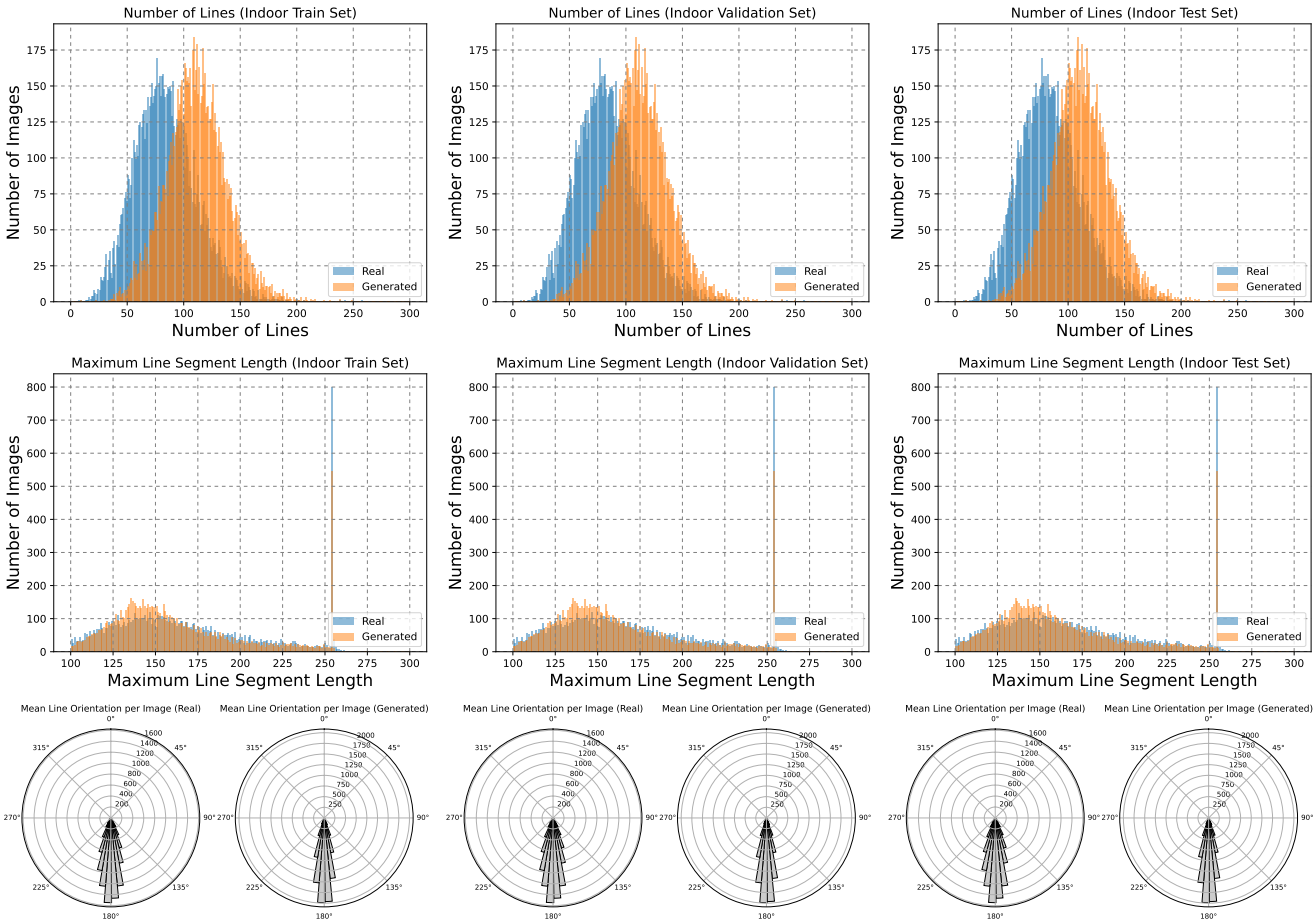


Figure 20. Line Segment Distribution in Indoor Scenes: We show the distribution of line segment counts and lengths in indoor scenes across training, validation, and test sets. The histograms (top row) compare the number of line segments detected in real versus generated images, with generated images generally exhibiting a different distribution, suggesting a discrepancy in line segment occurrence. The line segment length plots (middle row) show the maximum length of line segments. The polar plots (bottom row) illustrate the mean line orientation per image. While these basic statistical differences provide some discriminative power, they are notably less effective than our PointNet classifiers, which demonstrate a profound ability to detect and focus on critical geometric inconsistencies, as validated by our comprehensive ROC analysis in Figure 12.

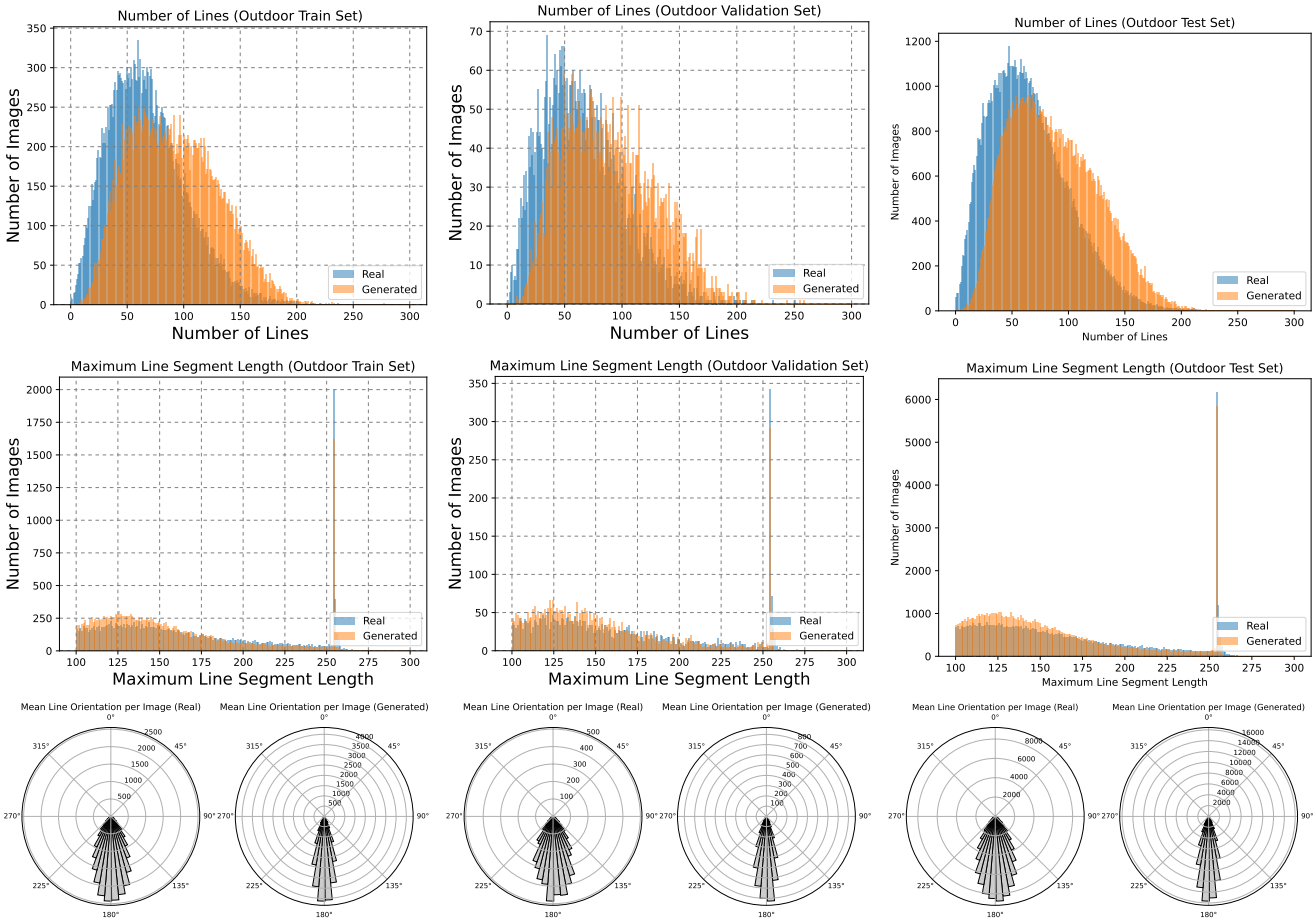


Figure 21. Line Segment Distribution in Outdoor Scenes: We show the distribution of line segment counts and lengths in indoor scenes across training, validation, and test sets. The histograms (top row) compare the number of line segments detected in real versus generated images, with generated images generally exhibiting a different distribution, suggesting a discrepancy in line segment occurrence. The line segment length plots (middle row) show the maximum length of line segments. The polar plots (bottom row) illustrate the mean line orientation per image. While these basic statistical differences provide some discriminative power, they are notably less effective than our PointNet classifiers, which demonstrate a profound ability to detect and focus on critical geometric inconsistencies, as validated by our comprehensive ROC analysis in Figure 12.