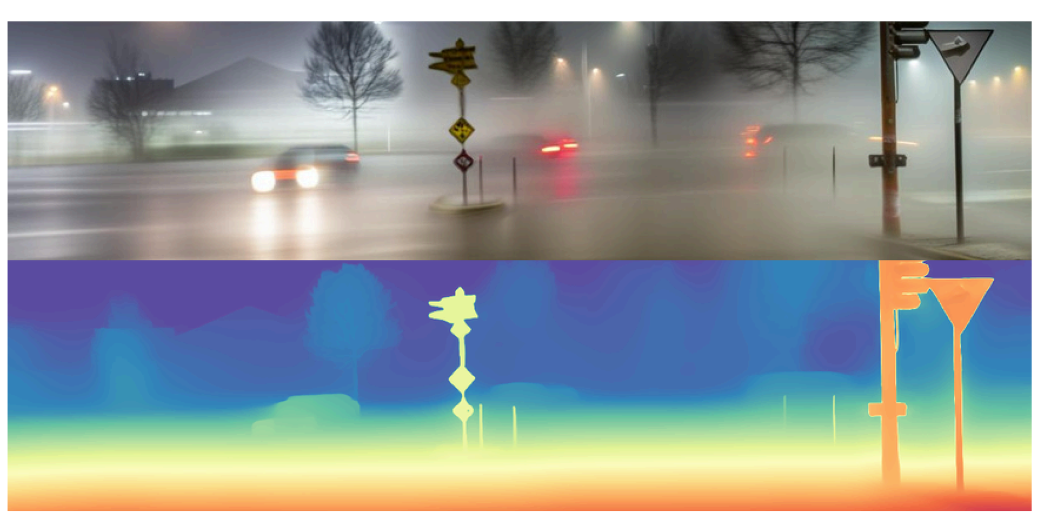
Blur-Aware Depth Estimation: Enhance Perception in Motion

Boyang Li, Kecen Yao, Jinyang Zhao
MScAC, Computer Science Department, University of Toronto

Motivation

- Depth estimation helps computers understand how far away things are in images, which is essential for various applications like autonomous driving, robotics, and augmented reality.
- Usually, to make depth estimation work well, we need a lot of data that shows different situations and conditions. However, getting this data can be hard, especially in tough situations like when it's very dark, raining, or in places with lots of movement. Most of the time, the data we have isn't varied enough, which means the systems we build don't work well in all situations—they can make mistakes when things look different than what they've been trained on.
- A well-known depth estimator, **Depth Anything v2,** has done a great job on most depth estimation scenarios. However, it struggles with motion blur. It tends to generate a depth map as if the blurred object is static, leading to inaccurate depth representations, especially for high-speed objects like cars, bicycles, or drones. In real-world scenarios, a blur-aware depth map would better represent the blurred object as it is perceived in motion, ensuring consistency and accuracy.



Related Work

In the paper "Diffusion Models for monocular depth estimation", authors used two advanced models to improve depth estimation in challenging conditions.

Firstly, they use **Depth-MiDaS** to estimate depth in simple, clear scenes and provide reliable depth information when conditions are not too difficult. Then, they used the **Stable-Diffusion Model** to generate realistic images by giving the original image and the estimated depth image from the first step. It can create lots of images under challenging conditions like rain, snow, or low light while keeping the positions of objects consistent.

Finally, they use those newly generated images and depth information to fine-tune the **Depth Everything V1** model, improving its ability to handle complex, real-world conditions.

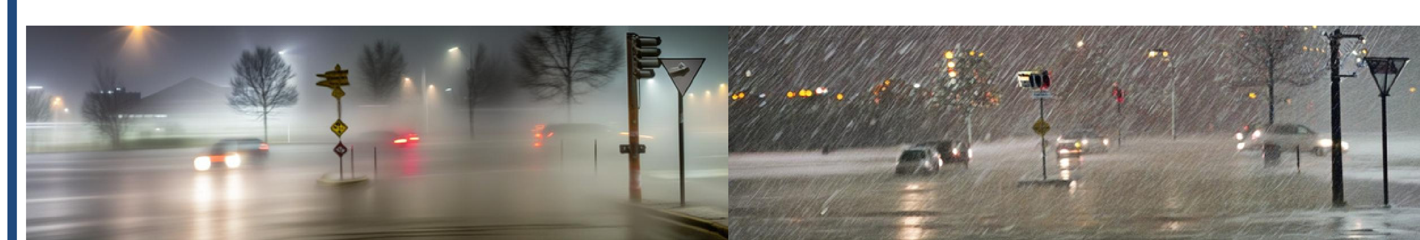
References

[1] Tosi, Zama Ramirez, and Poggi, "Diffusion Models for Monocular Depth Estimation: Overcoming Challenging Conditions," ECCV, 2024
[2] Yang et al., "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," CVPR, 2024
[3] Yang et al., "Depth Anything V2," NeurIPS, 2024

New Technique

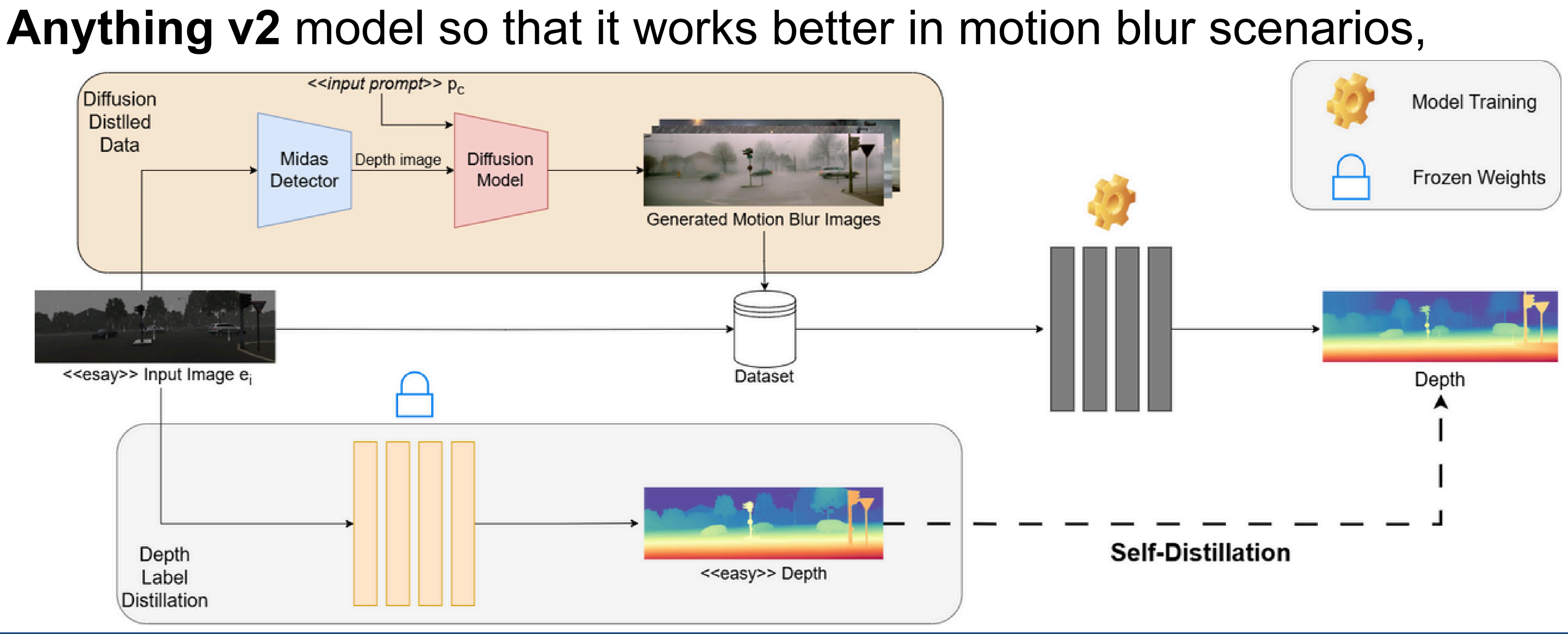
Our project uses the **stable diffusion** model with the **Depth-MiDaS** adapter to augment existing depth estimation datasets with synthetic but realistic RGB images for challenging environments. Based on the prompts used by Tosi et al. to generate images with adverse weather conditions, we add "motion blur" into each prompt together with a negative prompt containing "Static cars, sharp car details ..." in order to encourage diffusion model to start generating images with motion blur effect.

(a) RGB Image (b) Depth Image (c) "Hazardous smog blanketing city streets with motion blur"



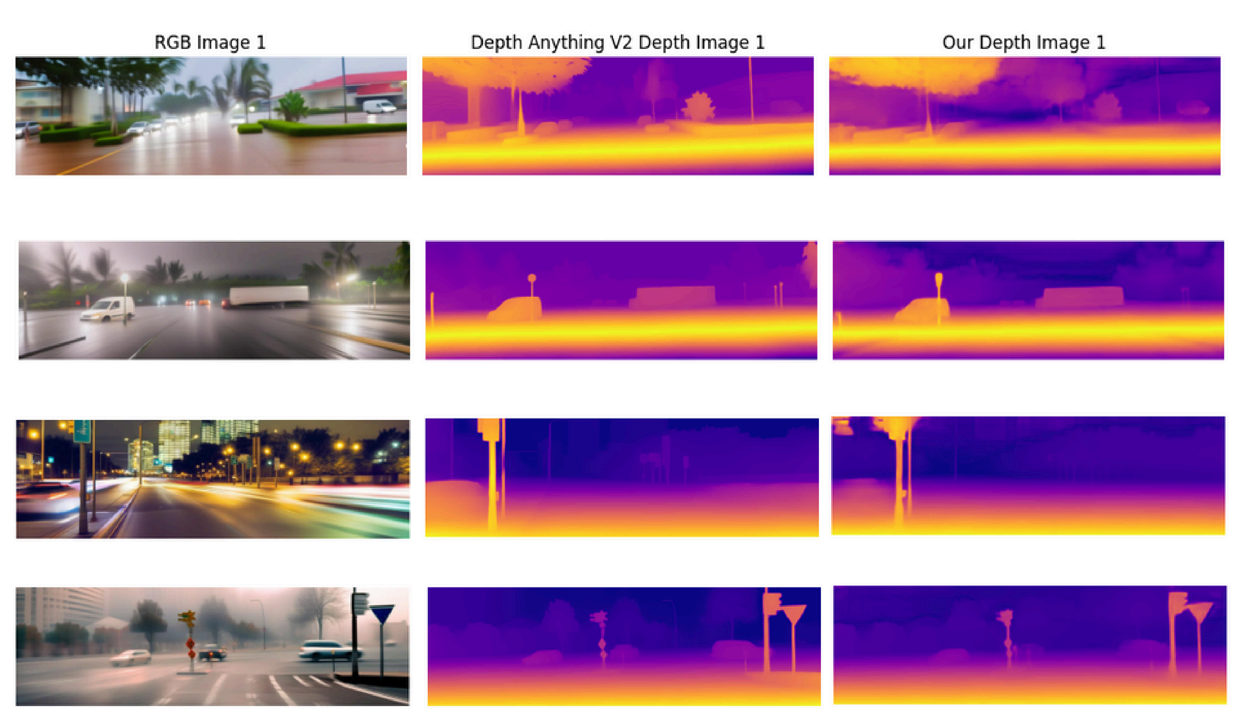
(d) "Driving with low visibility in urban fog with motion blur" (e) "Stranded in a car during intense hailstorm with motion blur" (f) "Urban roads with extreme sun glare with motion blur"

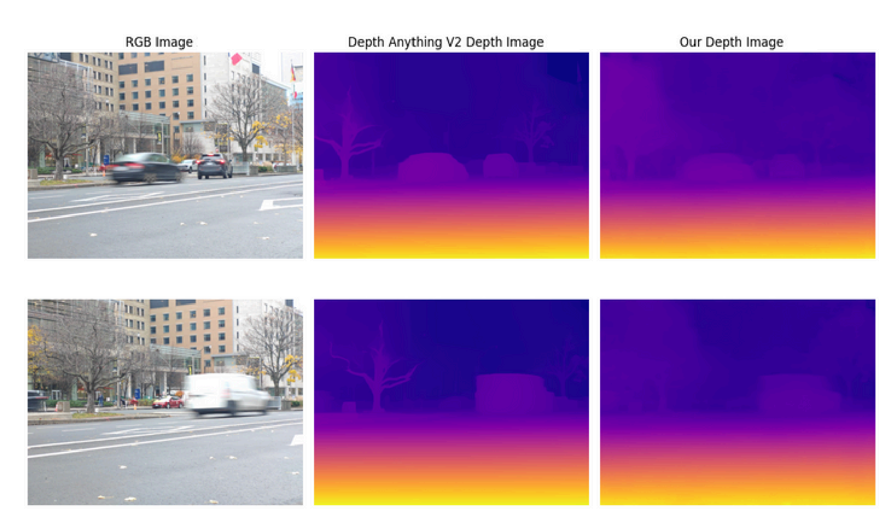
The following is the flow chart showing the method we have applied to fine-tune the **Depth Anything v2** model so that it works better in motion blur scenarios,



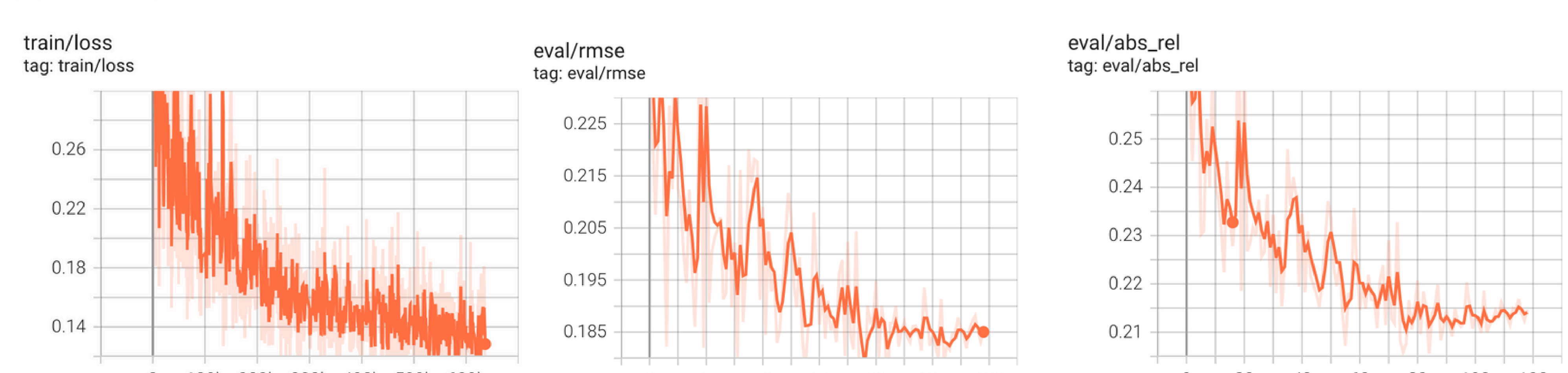
Experimental Results

Figure 1 (left) compares depth estimation under motion blur using RGB images from the KITTIv2 test set. Depth maps from the standard Depth Anything V2 model (middle column) are contrasted with those from our fine-tuned version (right column), optimized for motion blur. Figure 2 (right) extends this comparison to depth estimation on motion-blurred images captured by a mobile phone.





We can see that our fine-tuned model has the ability to incorporate motion blur into the depth field, thereby producing depth maps that faithfully represent both the static and dynamic elements of the scene. In contrast, the standard Depth Anything V2 model is limited to generating depth maps as if all objects were still, effectively ignoring the motion blur.



The training and validation metrics indicate effective fine-tuning of the model to handle motion blur in depth estimation. There are minor fluctuations in the early training stages, likely caused by variability in batch composition or the complexity of motion blur, stabilize over time. The concurrent reduction and stabilization of validation metrics confirm the model's robustness in integrating motion blur effects into depth estimation. Overall, the fine-tuned model achieves strong performance and effectively adapts to the motion blur domain.