Table 12. Inference speed comparison of DepthGen with DPT (Ranftl et al., 2021). Despite having an efficient denoiser backbone, DepthGen is much slower than DPT in total wall time due to the large number of denoising steps.

Method	Architecture	Resolution	Total Time [ms]	Inference steps	Per-step time [ms]
DPT-Hybrid	Nvidia RTX 2080	384×384	38	1	38
DepthGen	TPU v4	320×240	1089	128	8.5

Table 13. Performance of DepthGen on NYU with varying the number of denoising steps and a single sample. Fewer denoising steps can be used to achieve faster inference speed while sacrificing quality.

# steps	REL ↓	RMS ↓
128	0.075	0.324
64	0.077	0.324
32	0.086	0.342
24	0.104	0.378

B. Limitations

Diffusion models are known to be slow at inference since they need to run multiple denoising steps. This can be prohibitive for several vision tasks, including monocular depth estimation, where near real-time latency is often desired. Table 12 compares the inference speed of DepthGen with DPT (Ranftl et al., 2021). The statistics are reported for a model with 128 denoising steps since that is the default for the results in this paper. Despite having an efficient denoiser backbone, DepthGen is much slower than DPT in total wall time due to the large number of denoising steps. Hence, it is clear that the largest gains in inference latency can be achieved by reducing the number of denoising steps. Table 13 shows another interesting property of diffusion models; they allow trading off quality for inference speed by using a smaller number of denoising steps. DepthGen gives reasonable performance (compared to recent work) with as low as 24 denoising steps. However, a more thorough study into optimizing the inference speed of these models while preserving the generation quality is warranted. One potential direction for future work would be to explore the use of the recently proposed technique of progressive distillation (Salimans & Ho, 2022; Meng et al., 2022) which has been successfully used to distill generative image models with over 1000 denoising steps into those with 2-4 steps.