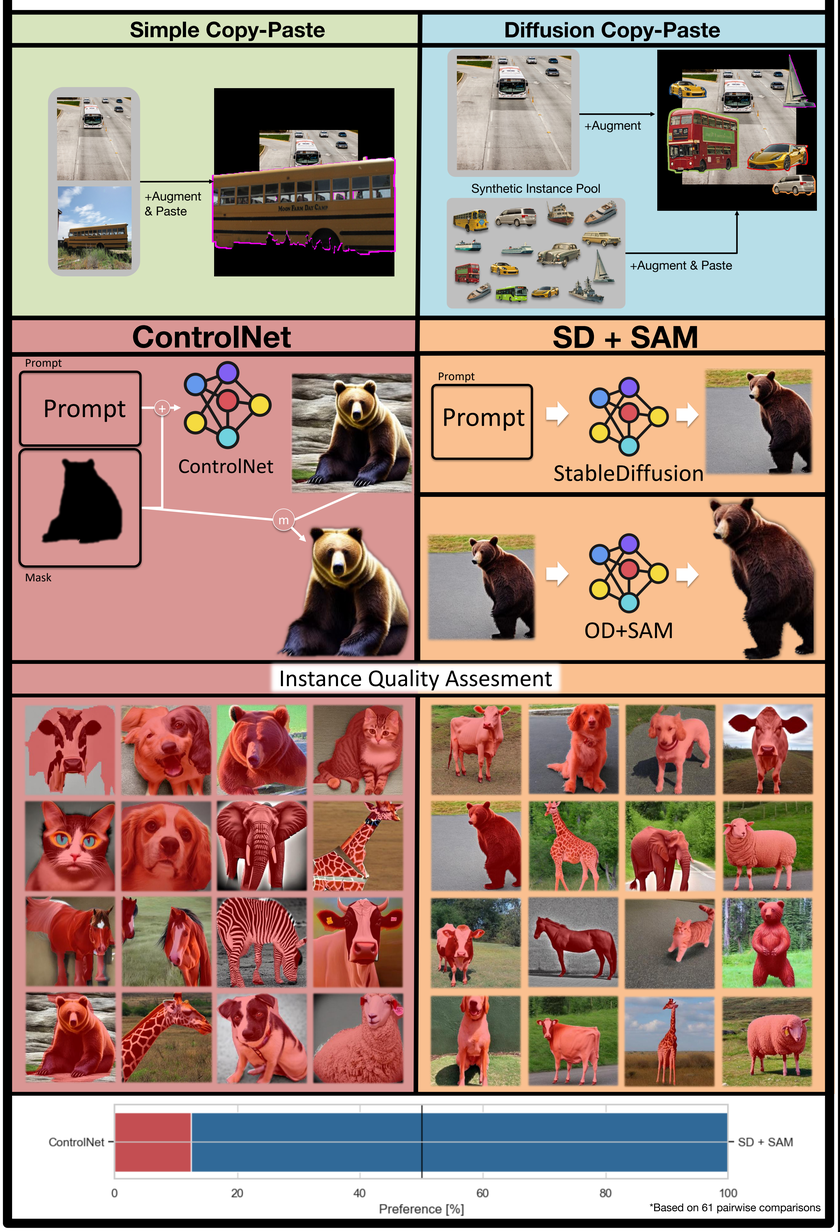
Improving Object Detection with Diffusion Augmentation

Erik Riise (s194633), Johannes Roed (s194648), Jens Parslov (S194649)

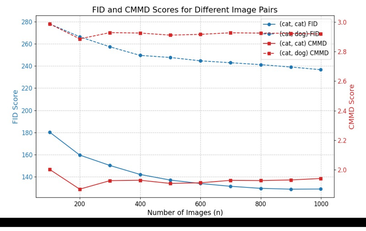
Advanved Deep Learning In Computer Vision
02501

Introduction



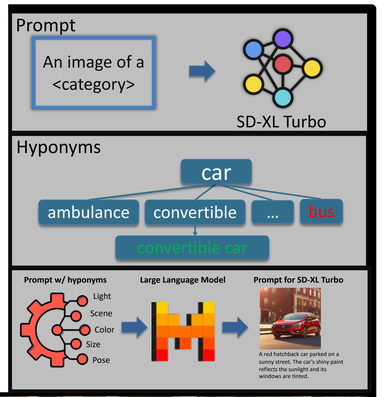
Metric: Assessing Generation

- We compare the generations using FID [1] and CMMD [2].
- FID and **CMMD** is calculated by masking out everything but the instance.
- FID and CMMD are calculated for multiple numbers (100, …, 1000) of images to check the stability of the metrics with limited data.



Generation Ablations

- Simple text template
- Hyponyms and synonyms from WordNet
- Manually remove misleading and modify ambigous
- We compare this method with using the class name
- We compare 1 vs 4 steps for SD-XL Turbo
- Experiment with LLM for added diversity
- Calculate CMMD and DINOv2 feature standard deviation



/wo Hyponyms + 1 step

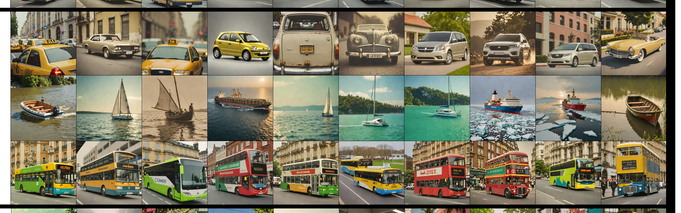
Not diverse
Not detailed

/wo Hyponyms + 4 steps

Less diverse
Detailed

/w Hyponyms + 4 steps

Diverse
Detailed

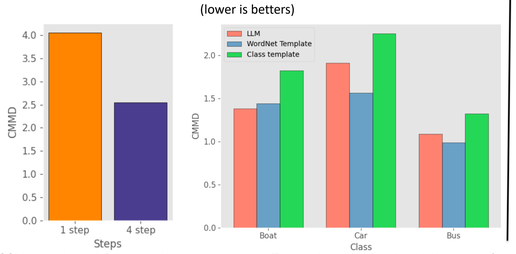


/w LLM + 4 steps

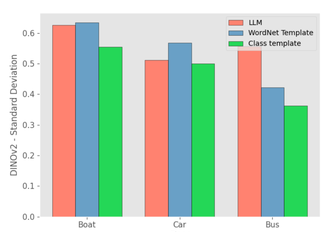
Diverse
Detailed

Comparing Distributions

(lower is betters)

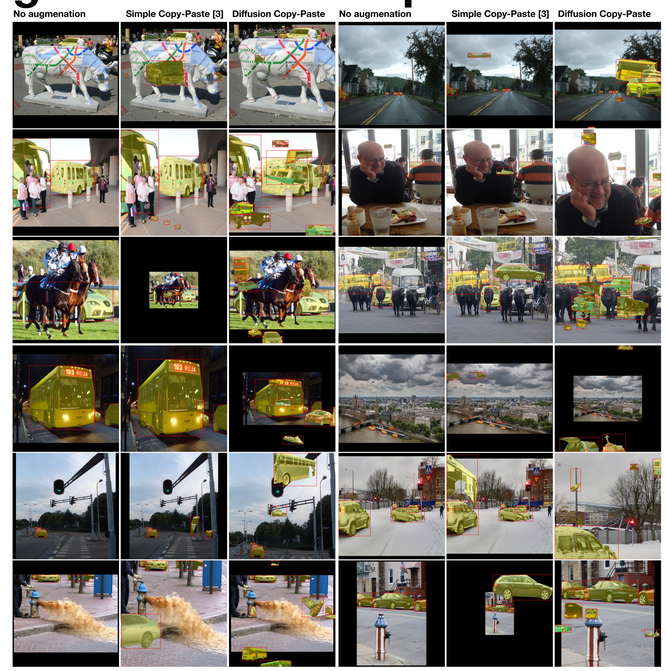


Diversity

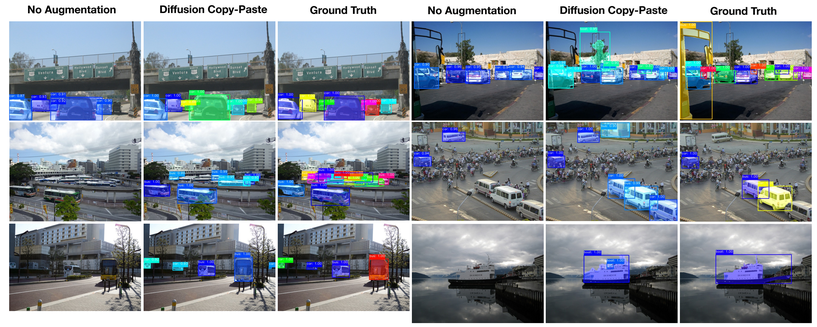


[1] Heusel, Martin & Ramsauer, Hubert & Unterthiner, Thomas & Nessler, Bernhard & Hochreiter, Sepp. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.
[2] Jayasumana, S., Ramalingam, S., Veit, A., Glasner, D., Chakrabarti, A., & Kumar, S. (2023). Rethinking fid: Towards a better evaluation metric for image generation. arXiv preprint arXiv:2401.09603.

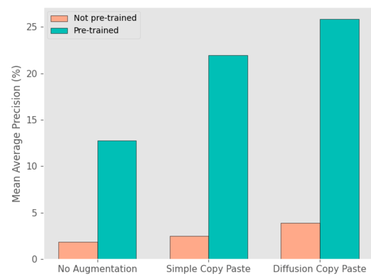
Augmentation Examples



Experiments with Mask-RCNN [4]



- Mask-RCNN [4]
- 420 training images with three classes
- Trained for 100 epochs, AdamW optimizer, 5 epoch linear lr warmup to 0.0003, followed by cosine annealing.
- /w & /wo pretrained backbone on ImageNet1k
- Box mAP evaluated on COCO validation set



[3] G. Ghiasi *et al.*, "Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 2917-2927, doi: 10.1109/CVPR46437.2021.00294.
[4] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.