DE-OCCLUSION OF OCCLUDED VEHICLE IMAGES FROM DRONE VIDEO

Urban Transport Systems Laboratory

Teacher: Pr. Geroliminis Nikolaos

Supervised by Tak Yura & Fonod Robert

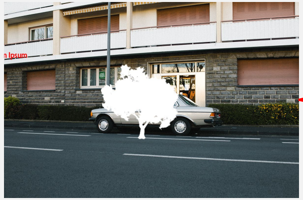
INTRODUCTION: DEFINITIONS

What does de-occlusion mean?

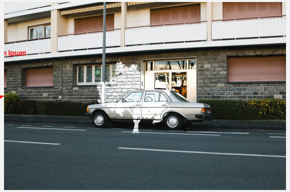
Segmentation of unwanted part

Remove that part

What is inpainting?



Generate part from surrounding



What is our goal?

Improve the detection capabilities of an UAV by removing the different obstacles from the targeted area.

LITERATURE REVIEW

PD-GAN: Probabilistic Diverse GAN for Image Inpainting :

A variant of the GAN model that uses more powerful probability properties to generates multiple and more meaningful outputs to inpaint missing regions.

No code available.

Visualizing the Invisible: Occluded Vehicle Segmentation and Recovery :

Use multiples GAN model with constraint applied to the mask segmentation. 3D model vehicles are used to constraints the masks.

No code available.

Large Scale Image Completion Via Co-Modulated Generative Adversarial Networks :

Allows the generator network to adaptively adjust its feature maps based on the conditioning vector, which enables it to capture high-level information about the image being generated.

No code available.

LITERATURE REVIEW

Image Completion with Heterogeneously Filtered Spectral Hints - SH GAN :

Introduce a new Spectral transform strategies:
Heterogeneous Filtering and Gaussian Split for large scale free form missing region inpainting.

Code available but with no training pipeline.

Can GAN Hallucinate Occluded People with a Plausible Aspect? :

Present the use of GANs for image enhancing in people attributes classification with occlusion.
Introduce an innovative way of creating a dataset with generated video games images

No code available.

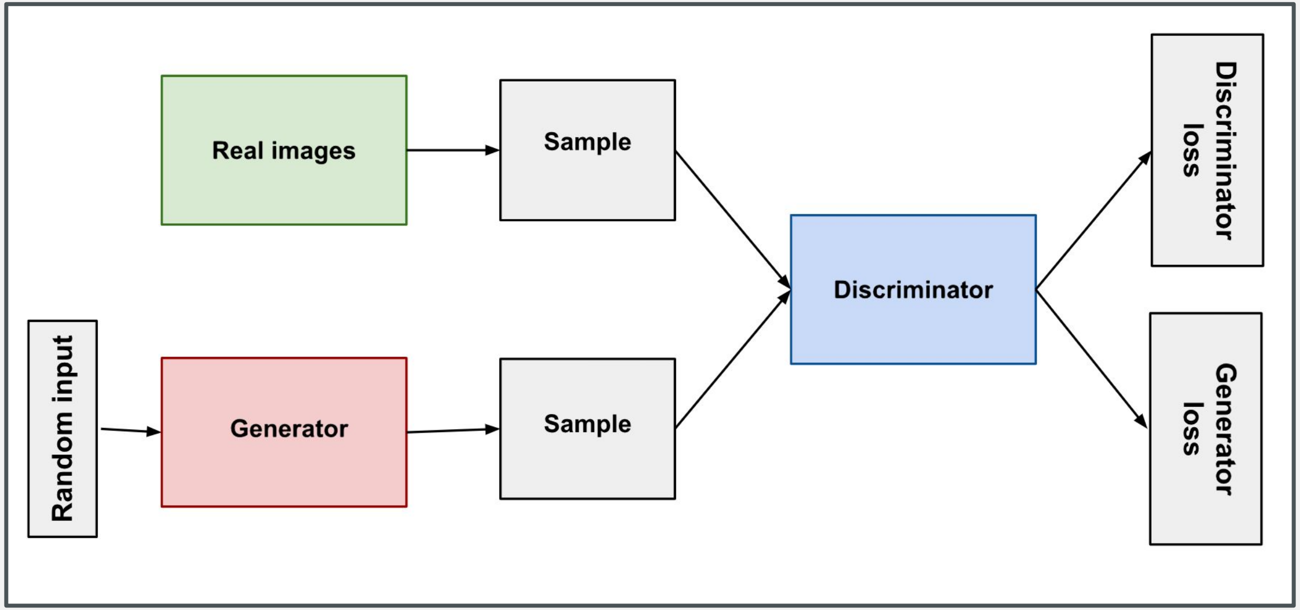
Complete Face Recovery GAN: Unsupervised Joint Face Rotation and De-Occlusion from a Single-View Image :

Very powerful Technique that uses the 3D Morphable Model (3DMM) a statistical model for faces shape and texture to create a 3D image with the correct shape and texture. By creating the 3D model, the model can deal with occlusions and face rotations. This technique is very powerful for face recognition.

Not applicable for vehicles because there is no model similar to 3DMM for vehicles..

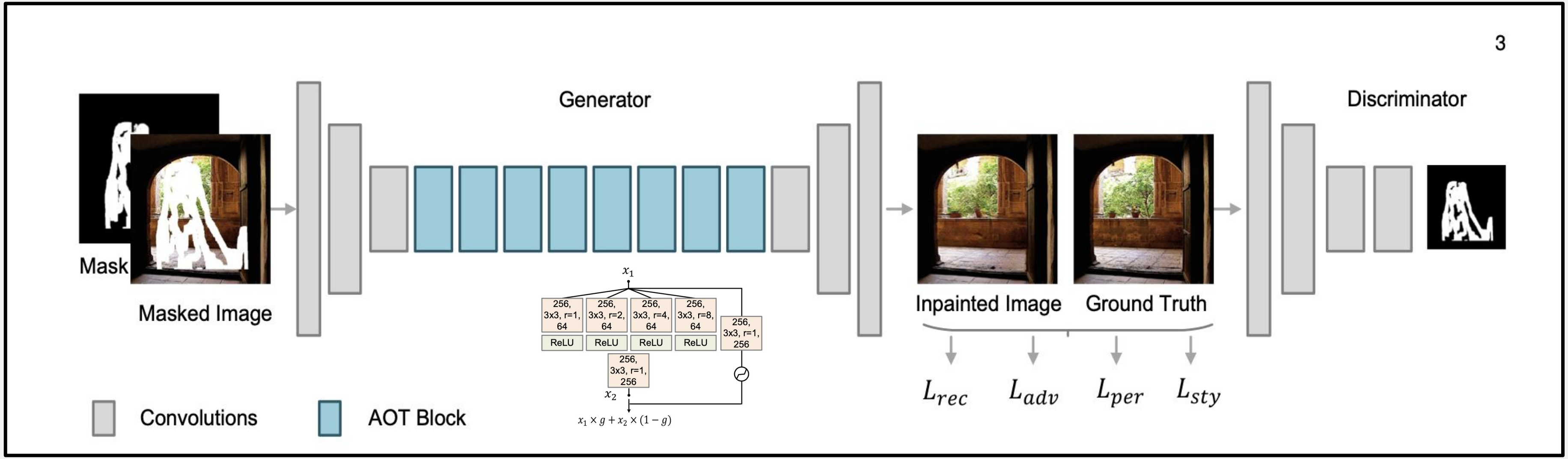
Generative Adversarial Network

What is a GAN?



AOT-GAN

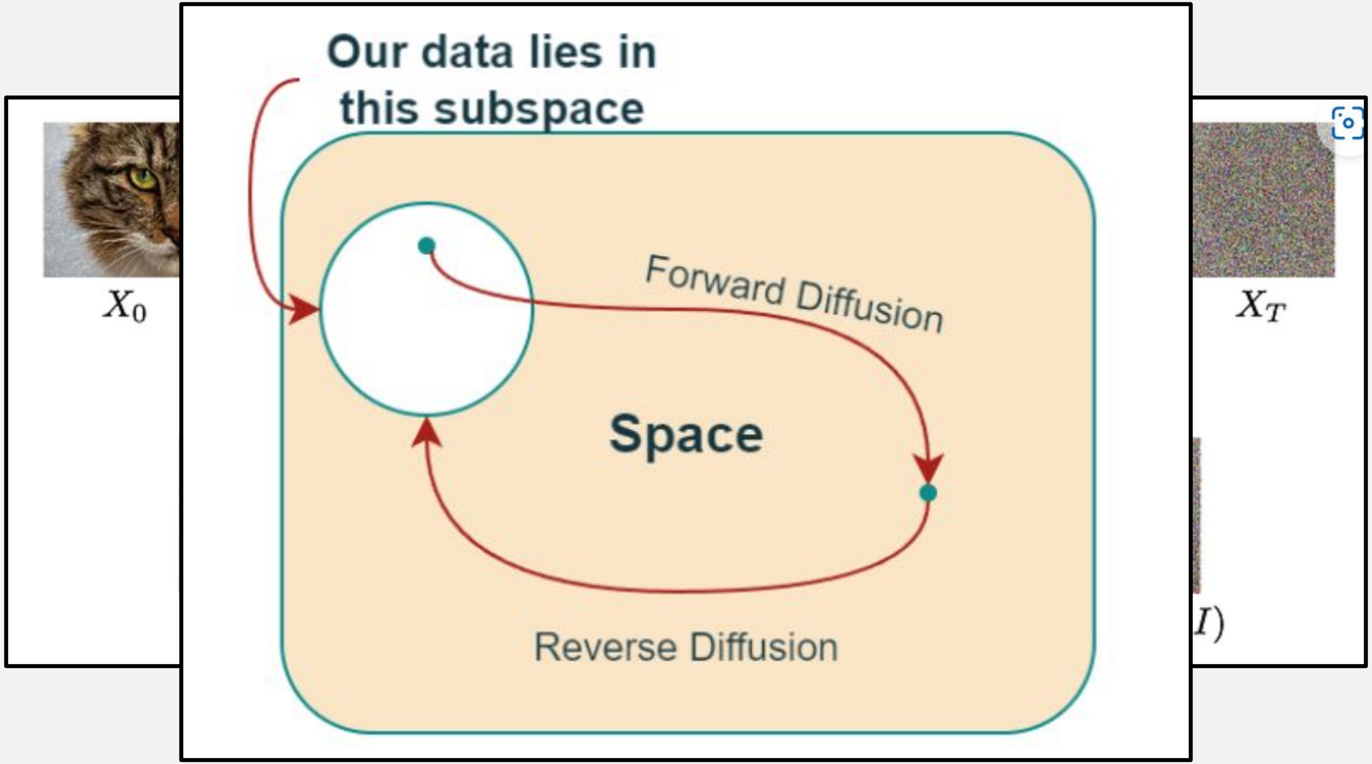
Aggregated Contextual Transformations for High-Resolution Image Inpainting

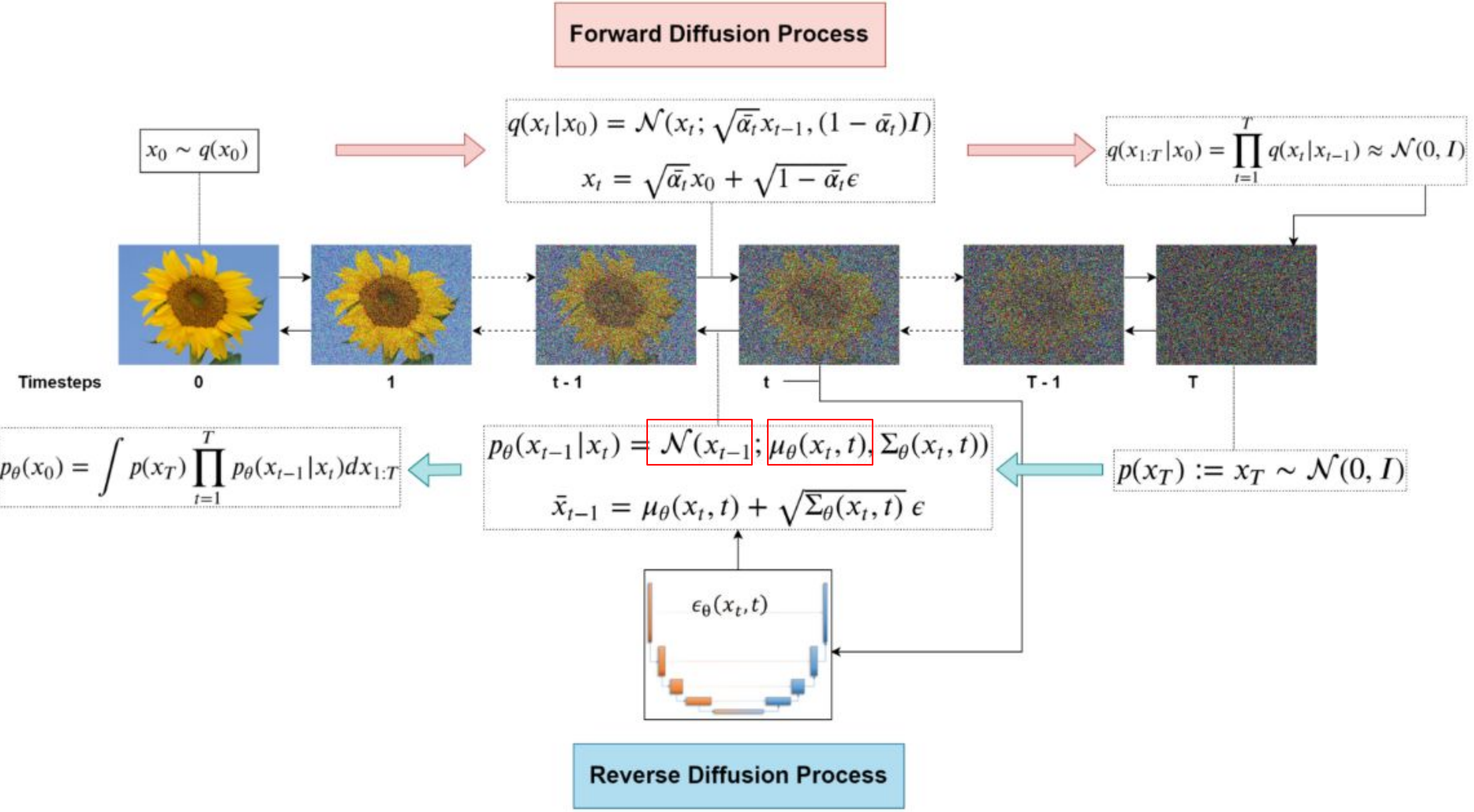


AOT-GAN : review

Good points (+)	Defaults (-)
Works with large images	More computational expensive than other GAN techniques
Works with very large masks	Cannot work on a set of image of different size.
Produces sharp, highly detailed and semantically meaningful images	
Harmonized generation	

Denoising Diffusion Probabilistic Models

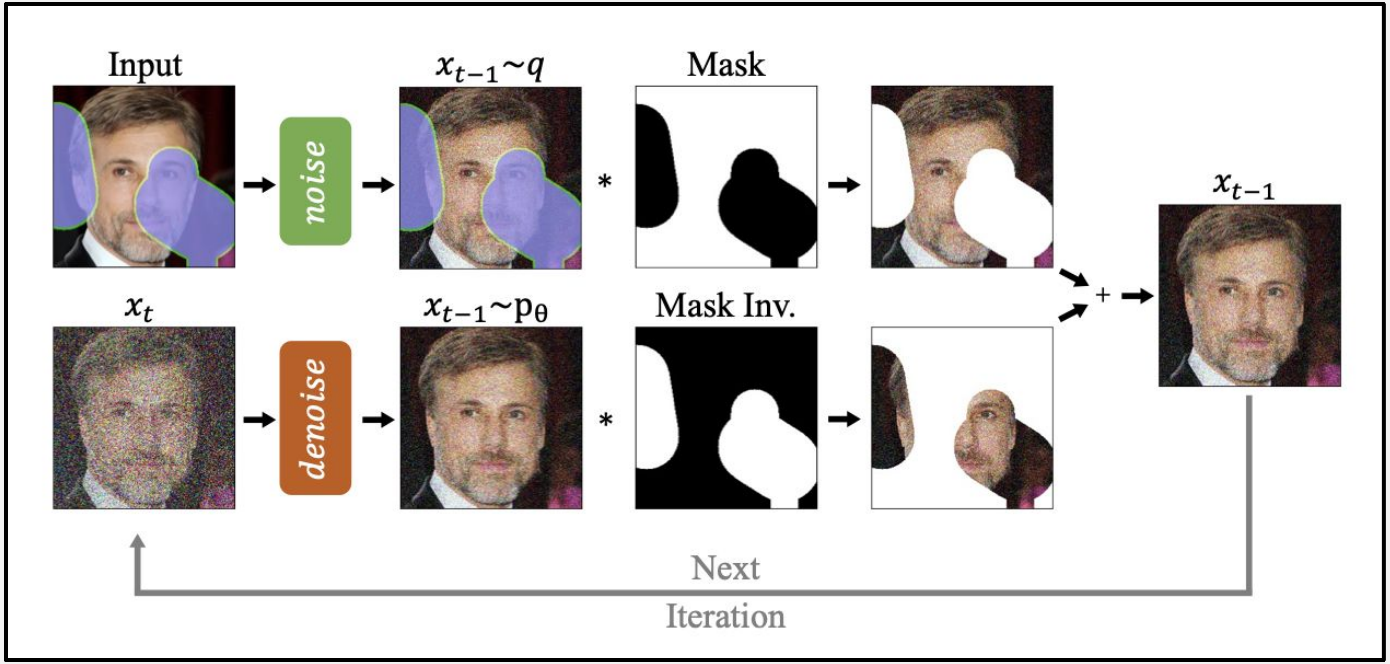




Forward Diffusion Process
$x_0 \sim q(x_0)$
$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_{t-1}, (1-\bar{\alpha}_t)I)$
$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$
$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}) \approx \mathcal{N}(0, I)$
Timesteps
0
1
t - 1
t
T - 1
T
$p_\theta(x_0) = \int p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)dx_{1:T}$
$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$
$\bar{x}_{t-1} = \mu_\theta(x_t, t) + \sqrt{\Sigma_\theta(x_t, t)}\,\epsilon$
$p(x_T) := x_T \sim \mathcal{N}(0, I)$
$\epsilon_\theta(x_t, t)$
Reverse Diffusion Process

REPAINT

RePaint: Inpainting using Denoising Diffusion Probabilistic Models



RePaint : review

Good points (+)	Defaults (-)
Requires no mask-specific training and generalize to any mask even big ones.	More computational expensive than other techniques. Sampling requires multiple steps, meaning that the generative process will be longer than it is for GANs or VAEs
Produces sharp, highly detailed and semantically meaningful images	Slow optimization process which makes it currently difficult to apply it for real-time applications
More flexible and diverse in the generation	Can produce realistic images completions that are very different from the Ground Truth image (Hallucination)
Harmonize generation (no border delimitation)	

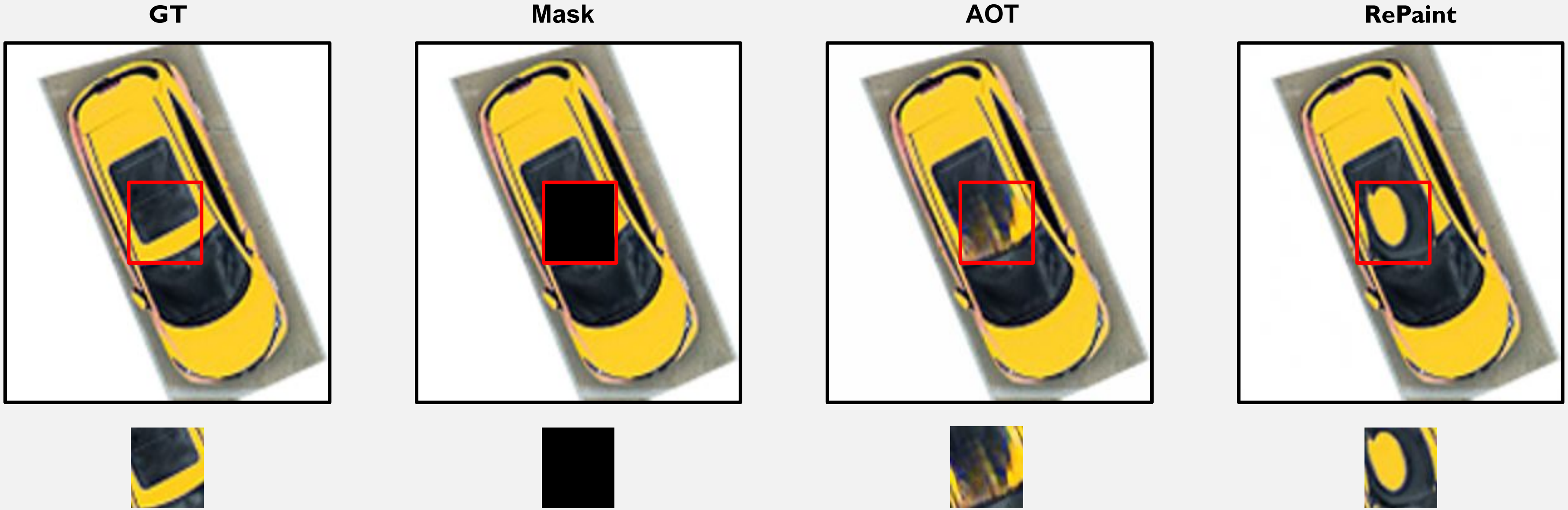
INTERMEDIATE RESULTS

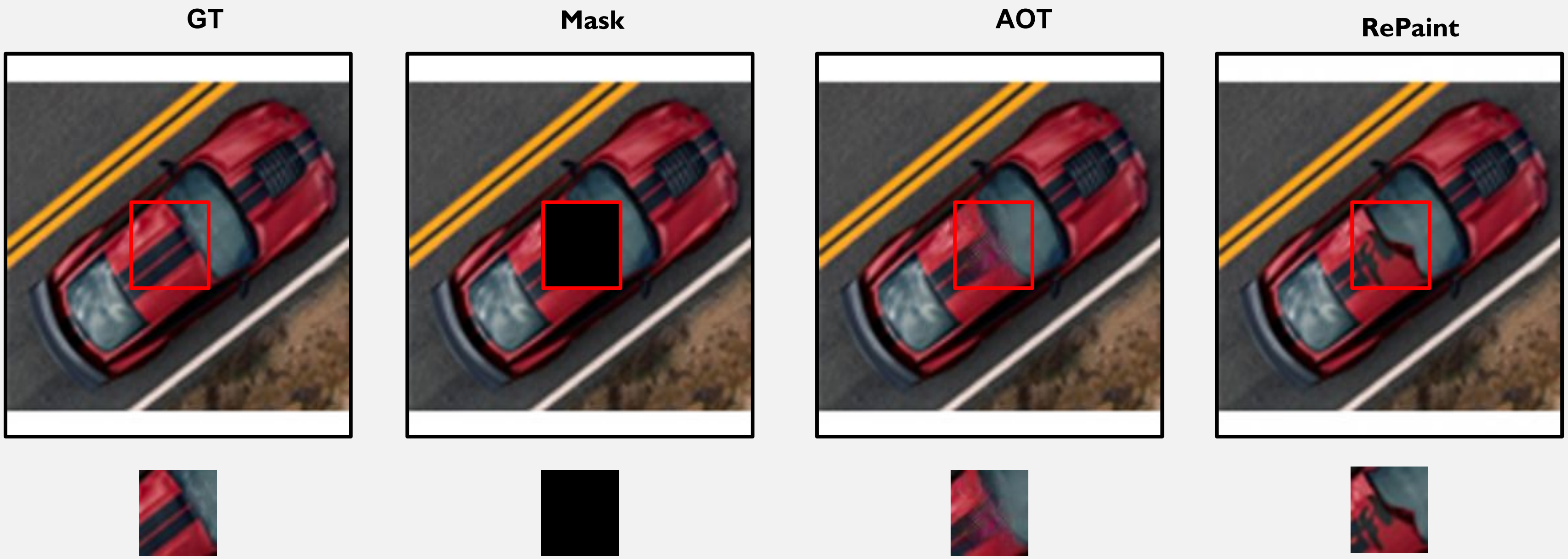
	GAN	Diffusion
Pros	Fast Sampling rate. High sample generation quality.	High sample generation quality. Diverse sample generation
Cons	Unstable training, low sample generation diversity (Mode Collapse)	Low sampling rate

Next are some results on the models respectively trained as on their release on their papers.
Image of 256x256 where used.

AOT GAN was trained on Places2 (building) and RePaint was trained on ImageNet (general but a lot of dog).

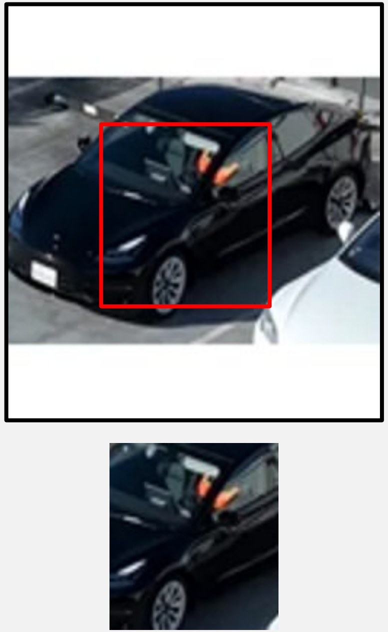
Small masks




GT
Mask
AOT
RePaint

Bigger Mask

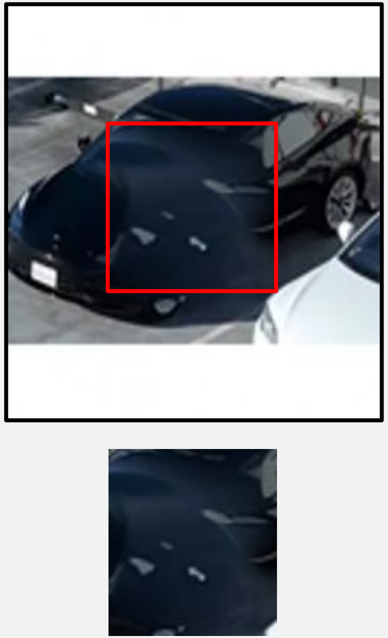
GT



Mask

AOT

Repaint




GT
Mask
AOT
RePaint

HOW TO MEASURE PERFORMANCE?

- **L1/Manhattan Norm** : Compute the mean absolute error between the generated image and the original one to determine the per-pixel reconstruction accuracy.
- **L2/Euclidean Norm** : Compute the mean squared error between the generated image and the original one to determine the per-pixel reconstruction accuracy.
- **SSIM** : It's a metric based on the human perception that outputs a number between 0 and 1 based on the similarity of three aspects : luminance, contrast & structure. An output of 1 means it's the same image.
- **PSNR** : It computes the ratio between the maximum possible power of the image and the power of the noise that is introduced during the reconstruction. The output is expressed as the logarithm of the PSNR ratio in db. A greater output means better performance.
- **LPIPS** : LPIPS is used to judge the perceptual similarity between two images. LPIPS essentially computes the similarity between the activations of two image patches for some pre-defined network. This measure has been shown to match human perception well. A low LPIPS score means that image patches are perceptual similar.
- **MAE** : measures the average magnitude of the errors in a set of predictions ie. the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.
- **FID** : The Fréchet distance compute the distance between two probability distributions : the distribution of the reconstructed images and the real images. A smaller output is a sign of good performance.

DIFFERENCES OF PERFORMANCE
Mid-semester

Mean of metrics on previous results

l1 -> 0
l2 -> 0
ssim -> 1
psnr -> inf
lpips -> 0.0
mae -> 0
fid -> 0

	AOT small	RP small	AOT big	RP big
l1	0.0906	0.1469	0.1364	0.1997
l2	0.0258	0.0683	0.046	0.086
ssim	0.7409	0.6573	0.531	0.4253
psnr	18.769	15.9563	14.2656	11.3369
lpips	2.3577	2.3411	2.7835	2.9692
mae	0.1316	0.2055	0.1895	0.2625
fid	337	322.99	311.92	316.922

DATASET CREATION

We created a script which permits the creation of our occluded vehicle datasets

Frequence of shape apparitions

input/output options

Mask and background color

```
%run main.py -in C:\Dataset\Pu\green -tro C:\Dataset\Pu/train_png -teo C:\Dataset\Pu/test_png
 --random 50 --circle 2 --ellipse 5 --poly 10 --rec 5 --star 5 --bgcol 0 0 0 --maskcol 255 255 255
 -r 0.7 --resize False --size 128 --hidden_ratio 0.5
```

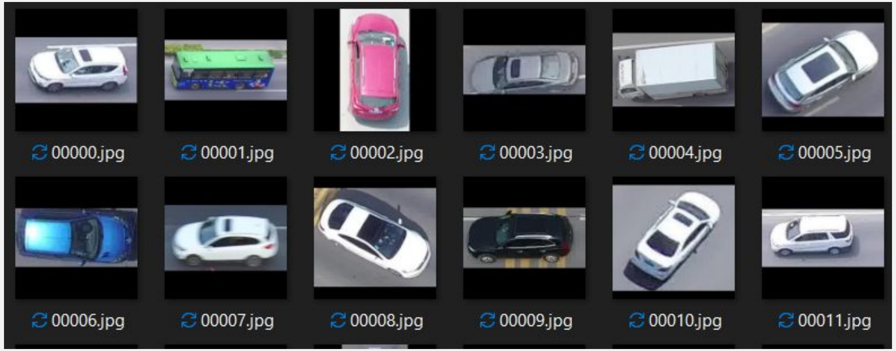
Ratio for test/train

Resize image

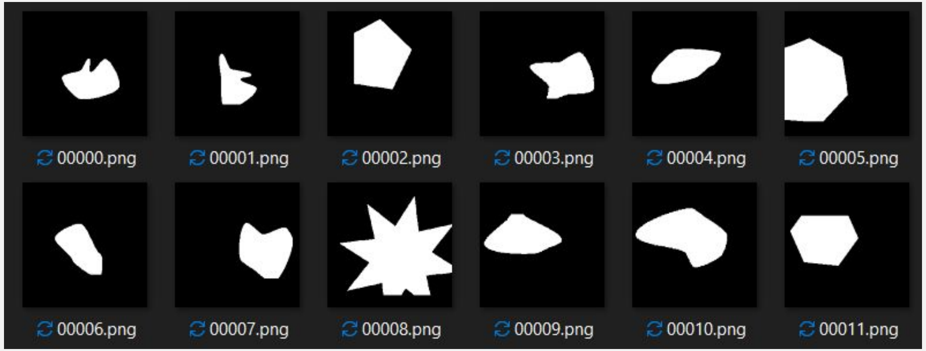
Size to aim for

How much percentage the mask must occlude the original image

Ground truth



Masks



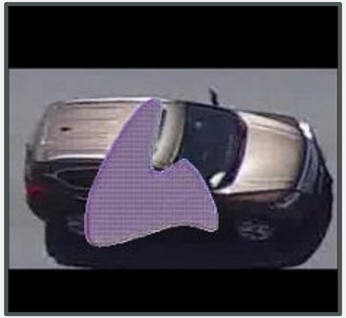
End semester results

Issue with the GAN

28.05.20203

01.06.20203

08.06.20203

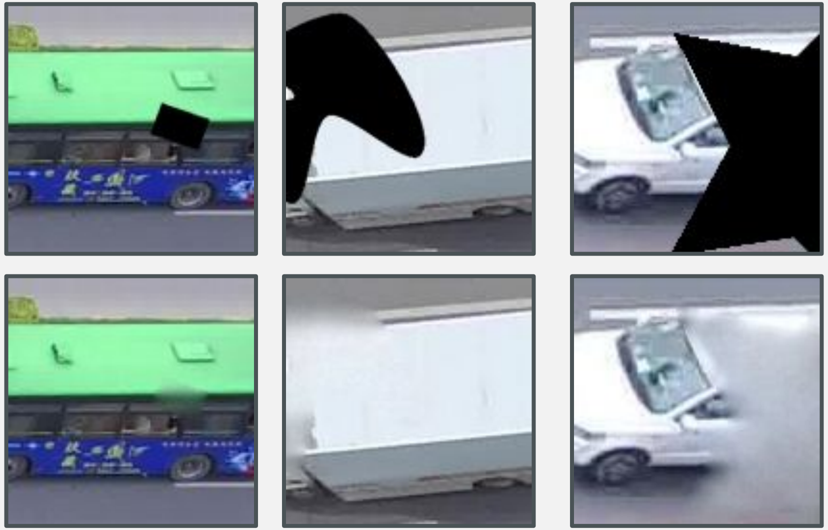


Issue Encountered with Repaint

```
Start VRAI
Traceback (most recent call last):
  File "/content/gdrive/MyDrive/Colab Notebooks/Repaint/test.py", line 180, in <module>
    main(conf_arg)
  File "/content/gdrive/MyDrive/Colab Notebooks/Repaint/test.py", line 69, in main
    model.load_state_dict(
  File "/usr/local/lib/python3.10/dist-packages/torch/nn/modules/module.py", line 2041, in load_state_dict
    raise RuntimeError('Error(s) in loading state_dict for {}:\n\t{}'.format(
RuntimeError: Error(s) in loading state_dict for UNetModel:
        Missing key(s) in state_dict: "input_blocks.3.0.in_layers.0.weight", "input_blocks.3.0.in_layers.0.bias", "input_blocks.3.0.in_layers.2.weight", "input_blocks.3.0.in_layers.2.
        Unexpected key(s) in state_dict: "input_blocks.3.0.op.weight", "input_blocks.3.0.op.bias", "input_blocks.6.0.op.weight", "input_blocks.6.0.op.bias", "input_blocks.9.0.op.weigh
```

Model incompatibilities between two different open source model

```
Evaluation Results
l1: 0.04634649679064751
l2: 0.012322023510932922
ssim: 0.8225930333137512
psnr: 22.066852569580078
lpips: 1.355625033378601
fid:
=====================================
```



CONCLUSION

- We provided a script to create occlusion dataset. And generated multiples one for vehicles (UAV POV)
- We successfully selected two of the newest model for inpainting and studied it
- We setup an mean (notebook) to evaluate the results with selected images metrics
- We setup two different pipelines to run/train Repaint and AOT GAN on Google Colab and partially on SCITAS.
- The GAN models trained with Places2 seems to outperform RePaint trained on ImageNet. But not really relevant as not trained on vehicle nor on the same dataset.

Thank you for your attention!