CZECH TECHNICAL UNIVERSITY IN PRAGUE

F3 Faculty of Electrical Engineering
Department of Cybernetics

Simulating depth measuring sensors for autonomous learning and benchmarking

Otakar Jašek
jasekota@fel.cvut.cz

Motivation

- Usage of the data from the state-of-the-art PC game (GTA V) *

---

* Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? CoRR, abs/1610.01983, 2016.

Motivation

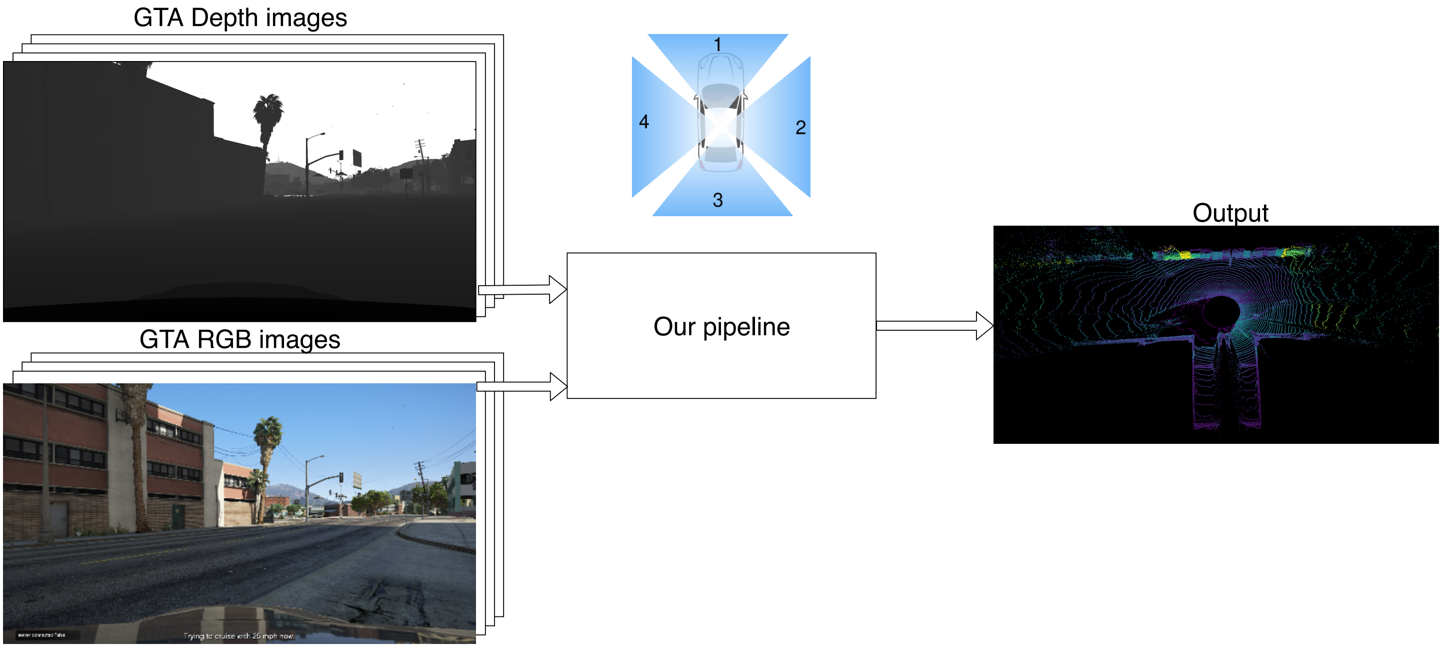
- Usage of the data from the state-of-the-art PC game (GTA V) *

- Only done for the visual tasks so far
- Try to model LiDAR sensors.

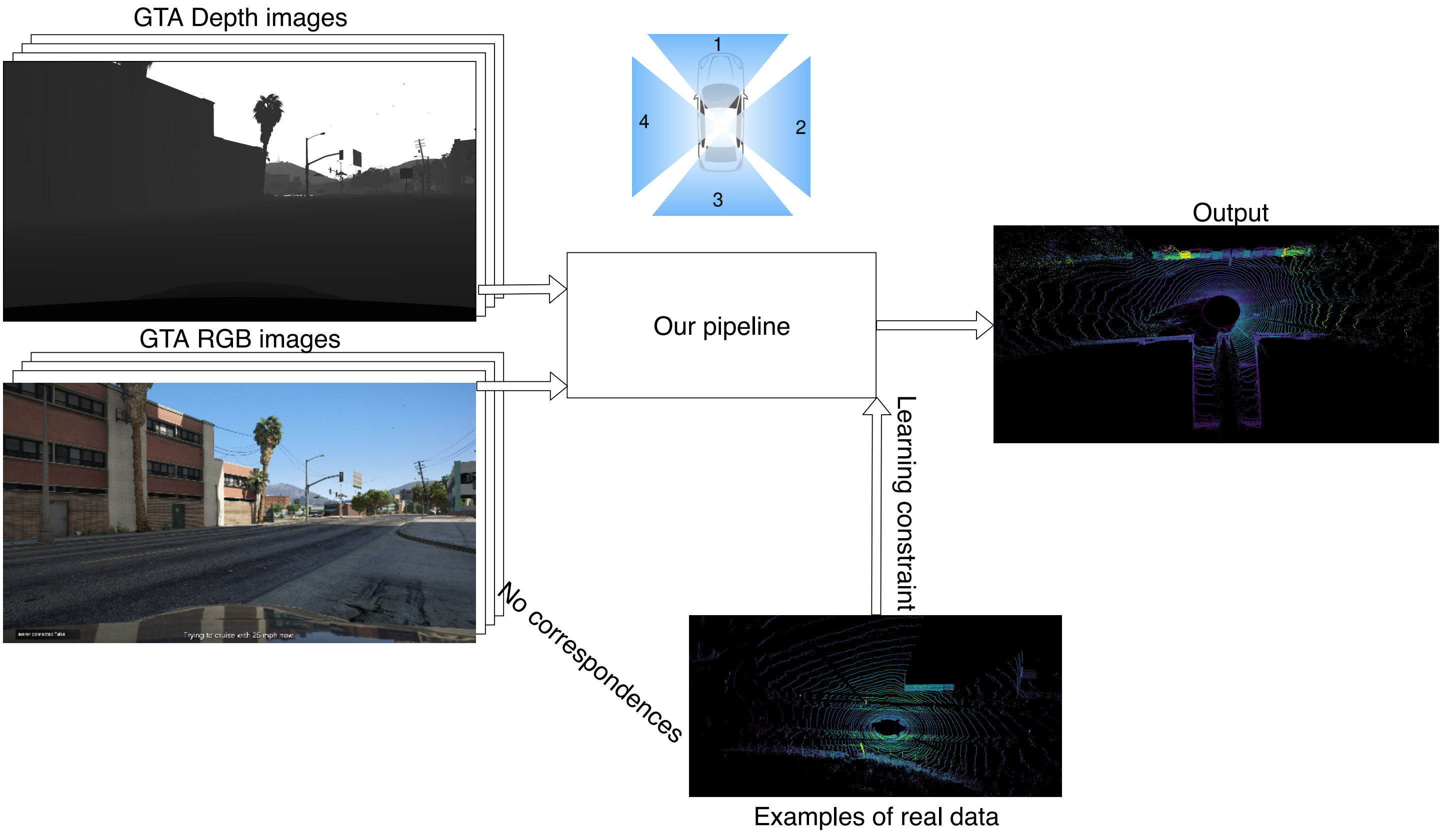
---

* Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? CoRR, abs/1610.01983, 2016.

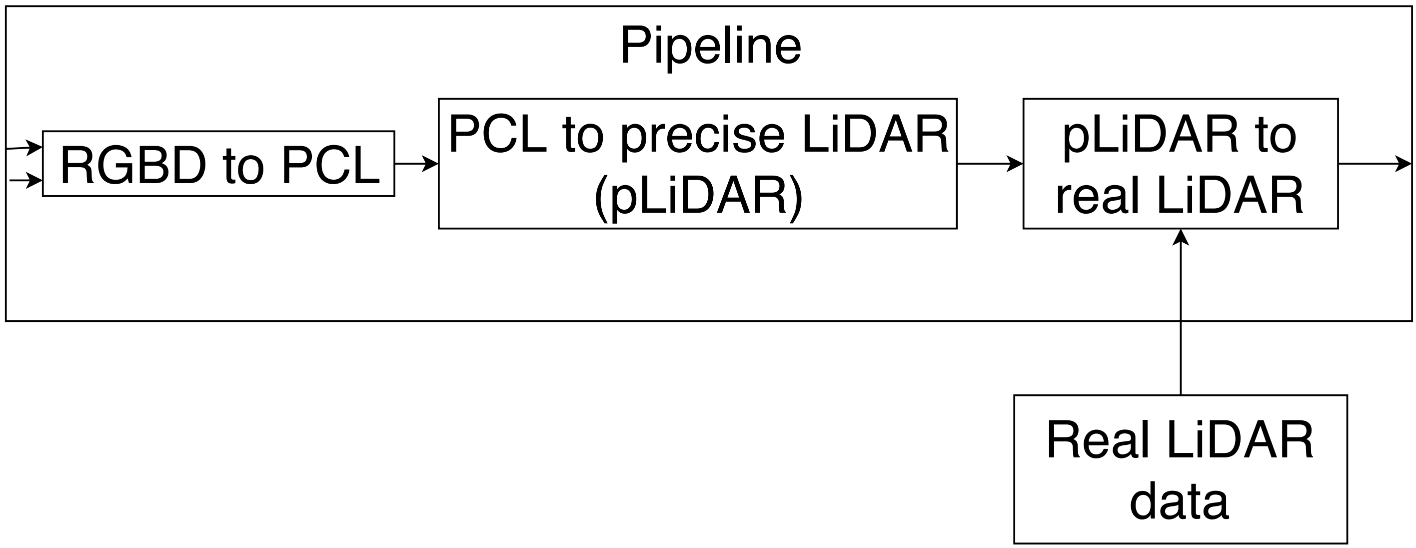
Goal



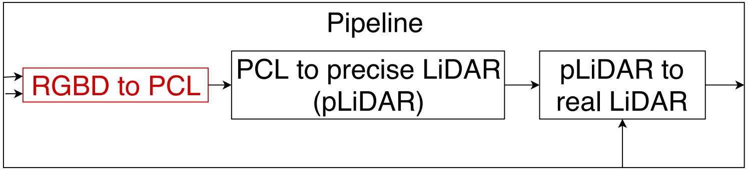
Goal


GTA Depth images
GTA RGB images
1
2
3
4
Our pipeline
Output
Learning constraint
No correspondences
Examples of real data

Goal

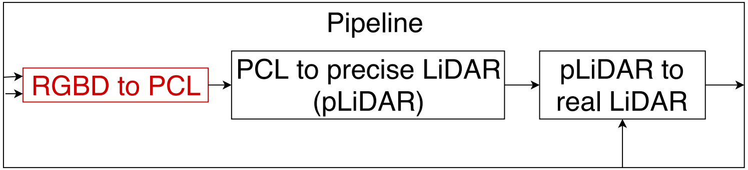

Pipeline
RGBD to PCL
PCL to precise LiDAR
(pLiDAR)
pLiDAR to
real LiDAR
Real LiDAR
data

RGBD to PCL

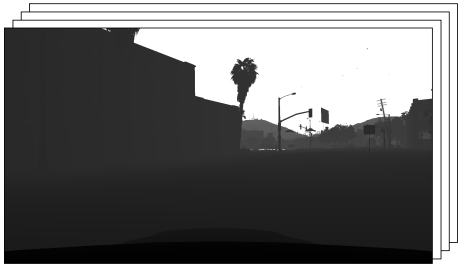


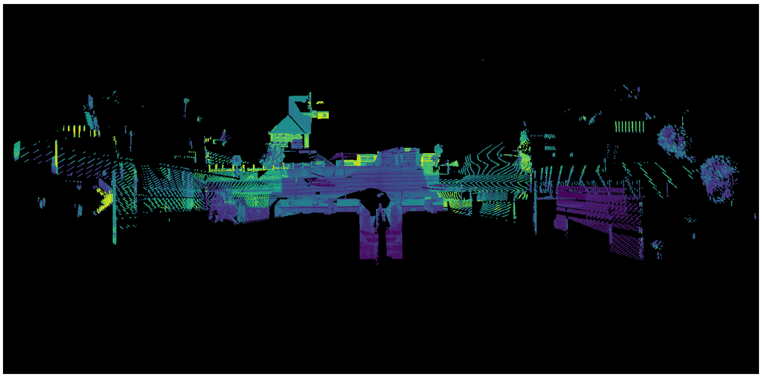
- Simple geometric task

RGBD to PCL

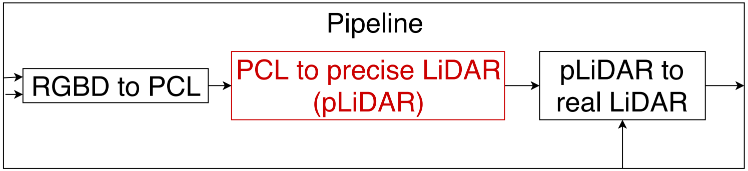


- Simple geometric task

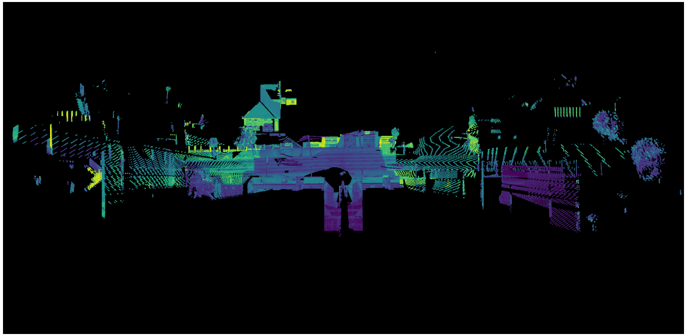




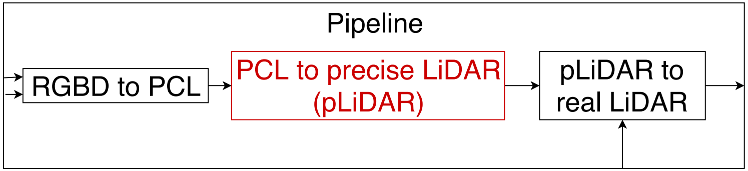
PCL to pLiDAR



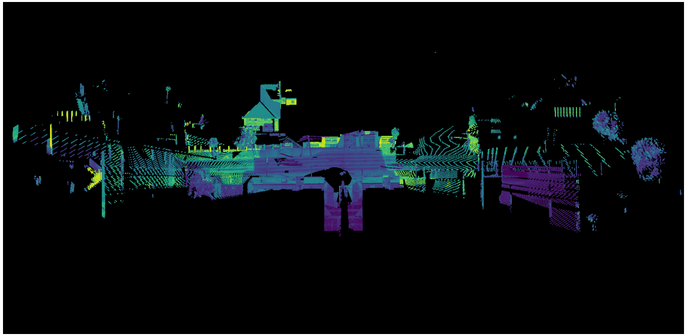
- Raycasting on 64×2084 grid

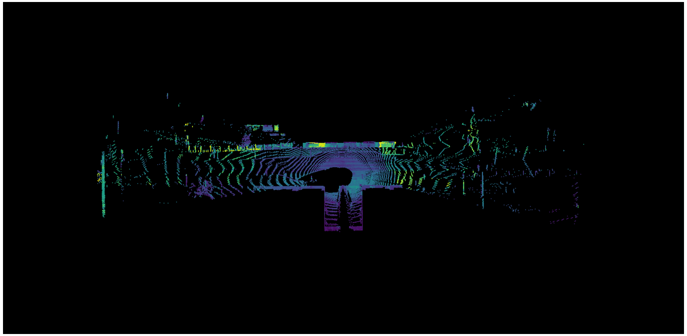


PCL to pLiDAR

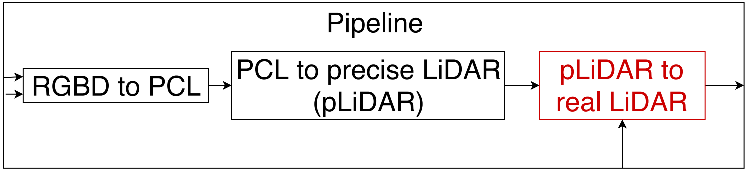


- Raycasting on 64 $\times$ 2084 grid

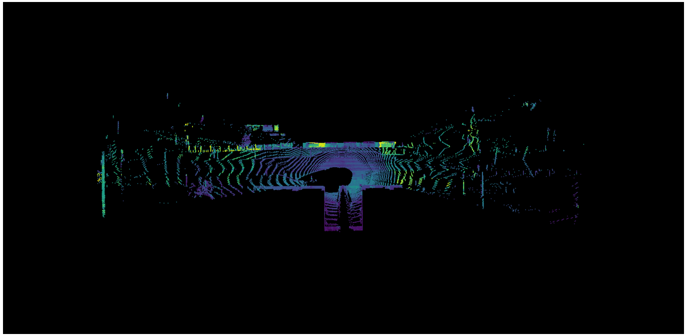




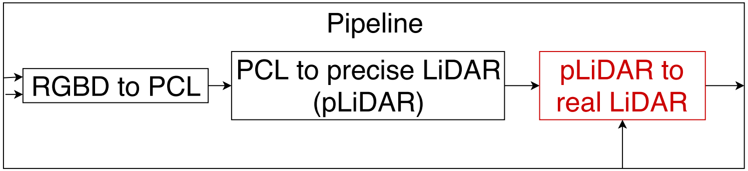
pLiDAR to real LiDAR



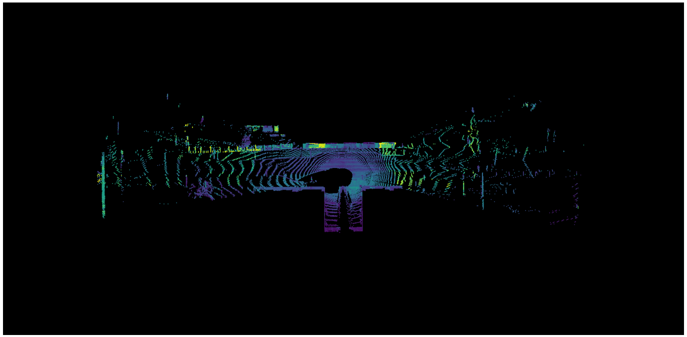
- Data driven approach using Generative adversarial networks (GANs)

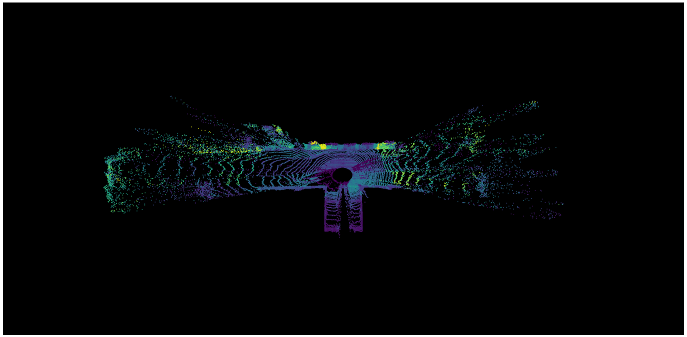


pLiDAR to real LiDAR



- **Data driven approach using Generative adversarial networks (GANs)**





GANs*

- Two neural networks – generator and discriminator.

* I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.

GANs*

- Two neural networks – generator and discriminator.
- Discriminator $D(\cdot)$ tries to recognize real from fake.
- Generator $G(\cdot)$ tries to fool discriminator.

* I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.

GANs*

- Two neural networks – generator and discriminator.
- Discriminator $D(\cdot)$ tries to recognize real from fake.
- Generator $G(\cdot)$ tries to fool discriminator.
- Therefore, it is two player minimax game.

$$\min_{G} \max_{D} V(G, D) = \mathbb{E}_{\mathbf{x} \sim p_{target}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{noise}}[\log(1 - D(G(\mathbf{z})))]$$

* I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.

GANs*

- Two neural networks – generator and discriminator.
- Discriminator $D(\cdot)$ tries to recognize real from fake.
- Generator $G(\cdot)$ tries to fool discriminator.
- Therefore, it is two player minimax game.

$$\min_G \max_D V(G, D) = \mathbb{E}_{\mathbf{x} \sim p_{target}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_{noise}}[\log(1 - D(G(\mathbf{z})))]$$

- Needed to train iteratively.
- $\mathcal{L}_G = \log(1 - D(G(\mathbf{z})))$
- $\mathcal{L}_D = -(\log D(\mathbf{x}) + \log(1 - D(G(\mathbf{z}))))$

* I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.

GAN variants

- 354 different GAN variants as of May 10th 2018
- They usually differ by used loss functions.

GAN variants

- 354 different GAN variants as of May 10th 2018
- They usually differ by used loss functions.

LSGAN

- $\mathcal{L}_G = \frac{1}{2}(1 - D(G(\mathbf{z})))^2$
- $\mathcal{L}_D = -\frac{1}{2}(D(\mathbf{x})^2 + (1 - D(G(\mathbf{z})))^2)$

GAN variants

- 354 different GAN variants as of May 10th 2018
- They usually differ by used loss functions.

LSGAN

- $\mathcal{L}_G = \frac{1}{2}(1 - D(G(\mathbf{z})))^2$
- $\mathcal{L}_D = -\frac{1}{2}(D(\mathbf{x})^2 + (1 - D(G(\mathbf{z})))^2)$

WGAN-GP

- $\mathcal{L}_G = -D(G(\mathbf{z}))$
- $\mathcal{L}_D = D(G(\mathbf{z})) - D(\mathbf{x}) + (||\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})||_2 - 1)^2$
- $\tilde{\mathbf{x}} = \epsilon \mathbf{x} + (1 - \epsilon) G(\mathbf{z})$
- $\epsilon \in [0, 1]$

GAN variants

- 354 different GAN variants as of May 10th 2018
- They usually differ by used loss functions.

LSGAN

- $\mathcal{L}_G = \frac{1}{2}(1 - D(G(\mathbf{z})))^2$
- $\mathcal{L}_D = -\frac{1}{2}(D(\mathbf{x})^2 + (1 - D(G(\mathbf{z})))^2)$

WGAN-GP

- $\mathcal{L}_G = -D(G(\mathbf{z}))$
- $\mathcal{L}_D = D(G(\mathbf{z})) - D(\mathbf{x}) + (||\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})||_2 - 1)^2$
- $\tilde{\mathbf{x}} = \epsilon \mathbf{x} + (1 - \epsilon) G(\mathbf{z})$
- $\epsilon \in [0, 1]$

SimGAN

- Refinement instead of generating.
- $\mathcal{L}_{reg} = ||\psi(\mathbf{x}) - \psi(G(\mathbf{x}))||_1$
- $\psi(\cdot)$ is a mapping preserving important information
- $\mathcal{L}_{reg}$ term added to generator.

CycleGAN

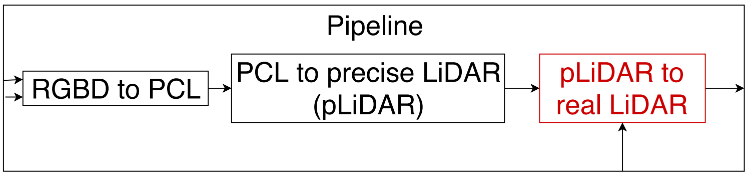
- Two GANs (4 networks together)
- $G_{X \to Y}$ and $G_{Y \to X}$
- The generators should be inverse mappings with respect to each other.

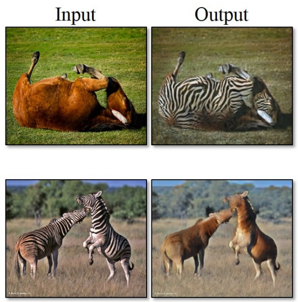
CycleGAN

- Two GANs (4 networks together)
- $G_{X\to Y}$ and $G_{Y\to X}$
- The generators should be inverse mappings with respect to each other.
- $\mathcal{L}_{cyc} = (G_{X\to Y}(G_{Y\to X}(\mathbf{y})) - \mathbf{y}) + (G_{Y\to X}(G_{X\to Y}(\mathbf{x})) - \mathbf{x})$
- This loss term is added to both generators.

CycleGAN

- Two GANs (4 networks together)
- $G_{X \to Y}$ and $G_{Y \to X}$
- The generators should be inverse mappings with respect to each other.
- $\mathcal{L}_{cyc} = (G_{X \to Y}(G_{Y \to X}(\mathbf{y})) - \mathbf{y}) + (G_{Y \to X}(G_{X \to Y}(\mathbf{x})) - \mathbf{x})$
- This loss term is added to both generators.





*

* Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Computer Vision (ICCV), 2017 IEEE International Conference on, 2017.

Experiments' setup

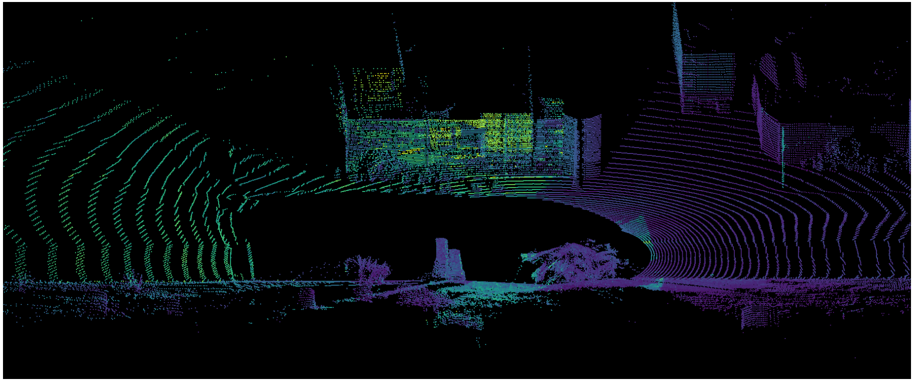
- Using CycleGAN to find mappings between real and GTA LiDAR data.
- Six different setups varying by the used loss functions of underlying GANs.
- Generator and discriminator stayed the same.

Experiments' setup

- Using CycleGAN to find mappings between real and GTA LiDAR data.
- Six different setups varying by the used loss functions of underlying GANs.
- Generator and discriminator stayed the same.
- Networks were rather shallow.
- Discriminator had:
  - $\sim$266k trainable parameters
  - 4 convolutional layers and one fully-connected
- Generator had:
  - $\sim$1.1M trainable parameters
  - 6 convolutional layers and 6 ResNet layers of size 2

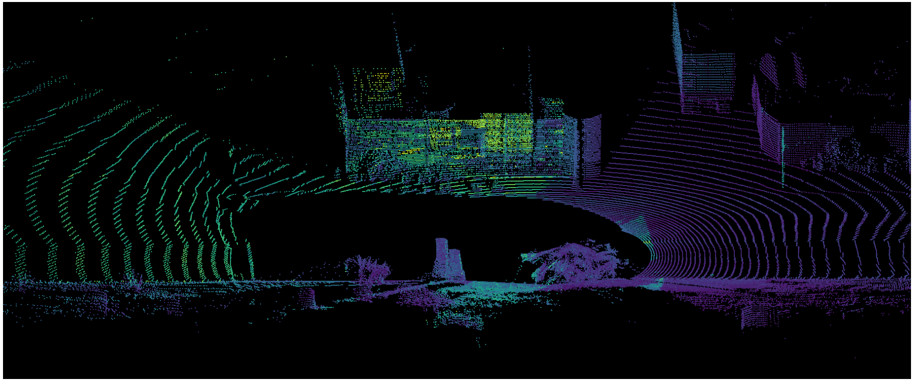
Experiments' results

GTA pLiDAR

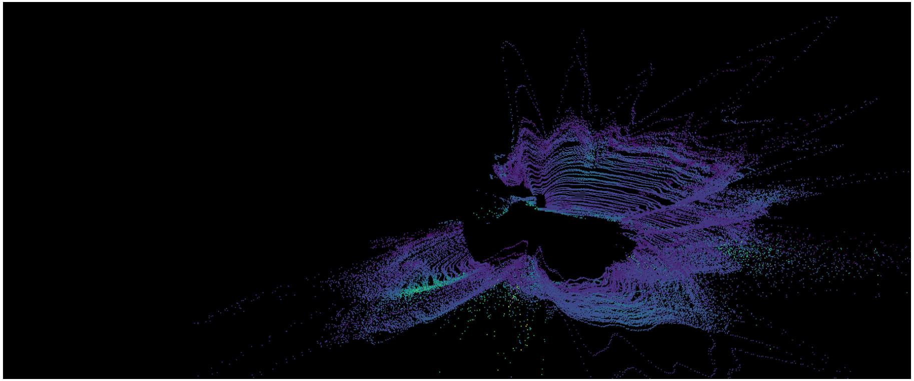


Experiments' results

GTA pLiDAR

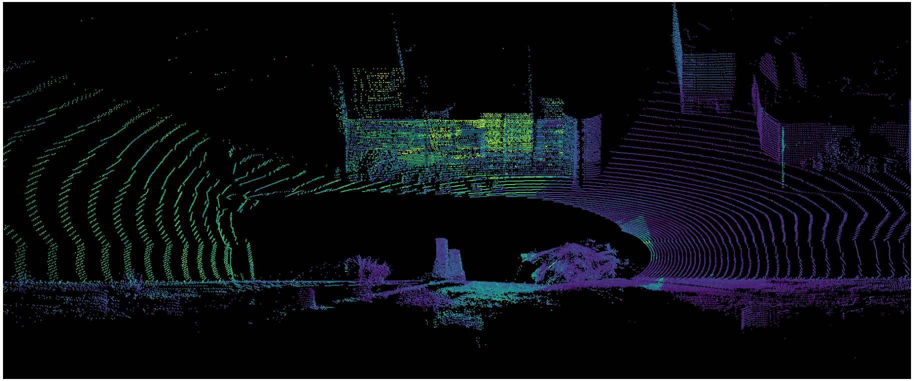


Original GAN

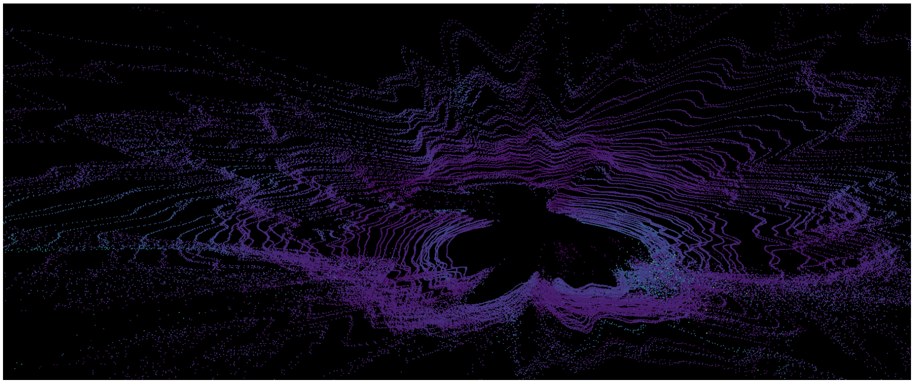


Experiments' results

GTA pLiDAR

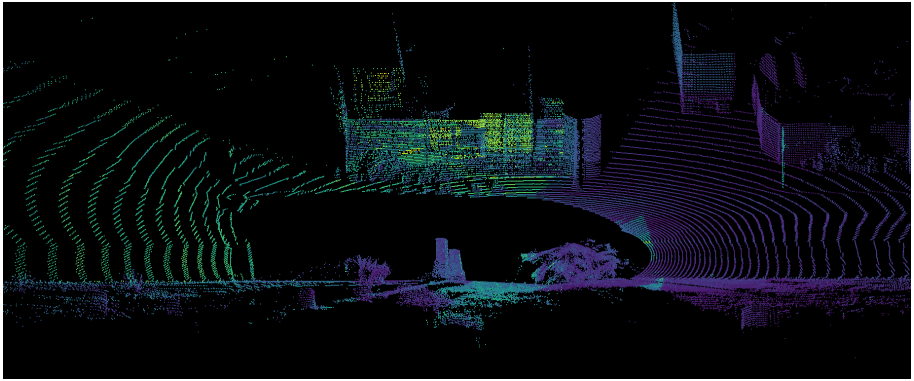


Original GAN + $\mathcal{L}_{reg}$

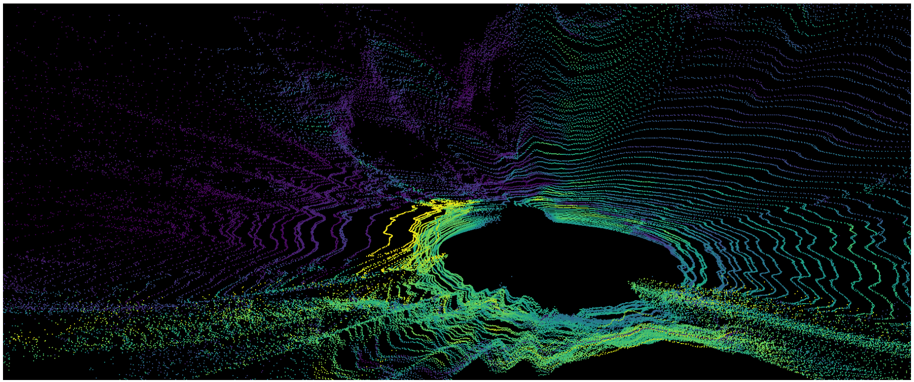


Experiments' results

GTA pLiDAR

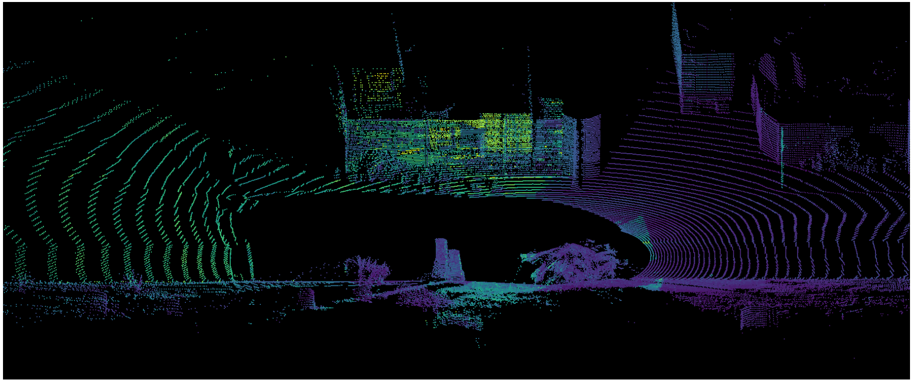


LSGAN

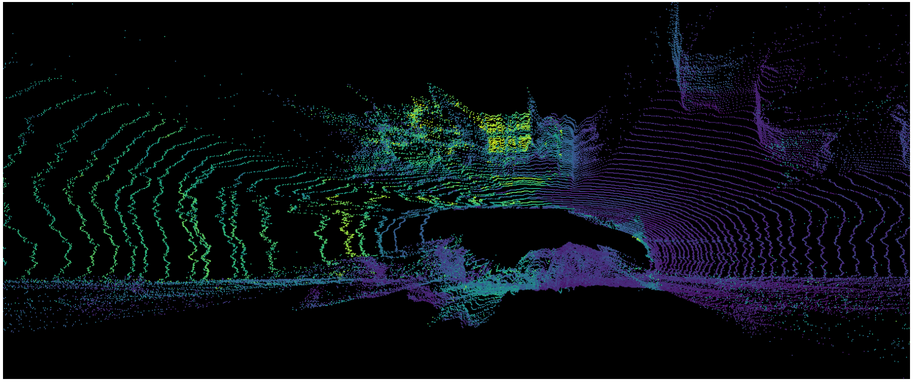


Experiments' results

GTA pLiDAR

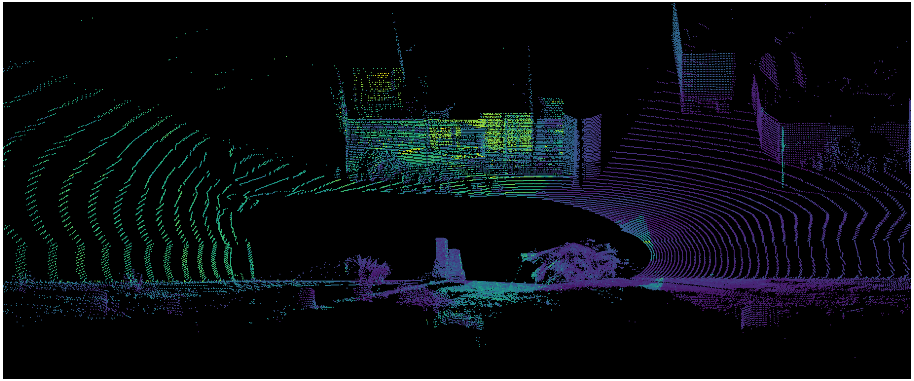


LSGAN + $\mathcal{L}_{reg}$

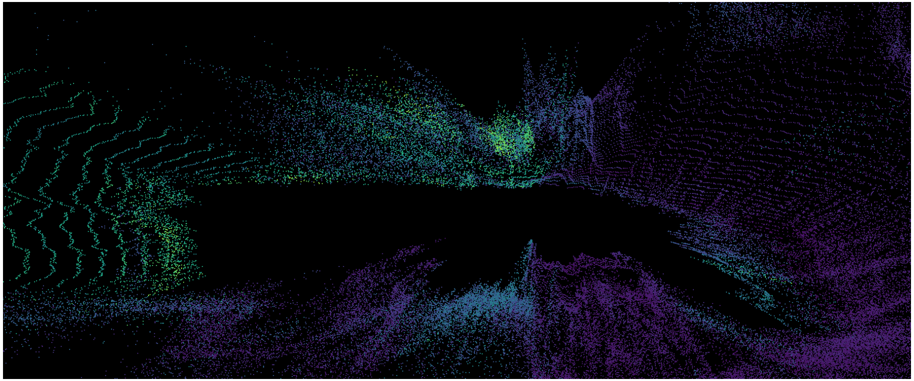


Experiments' results

GTA pLiDAR

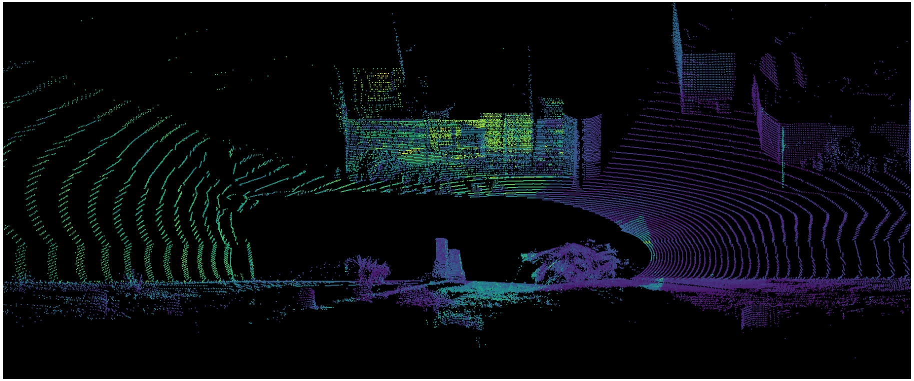


WGAN-GP

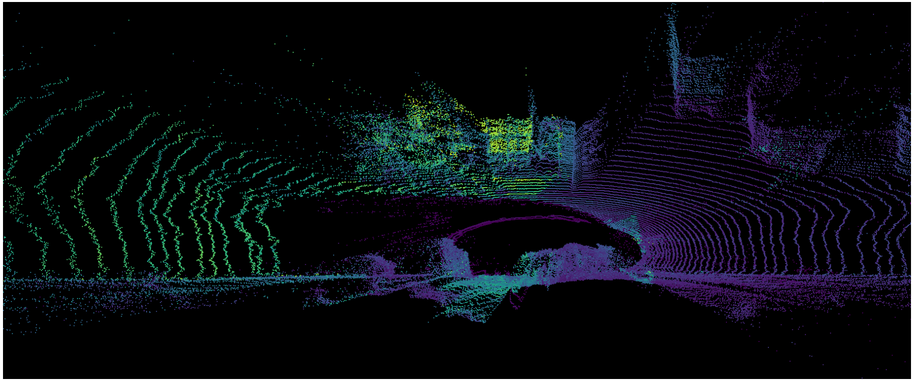


Experiments' results

GTA pLiDAR



WGAN-GP + $\mathcal{L}_{reg}$



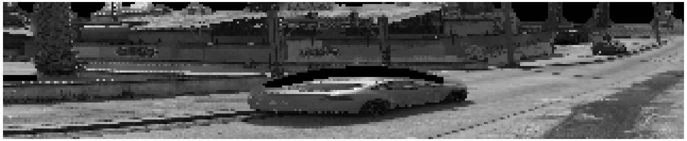
Evaluation

- No sensible metric of “success”
- However, the transformation is probably heading into a direction of the correct solution

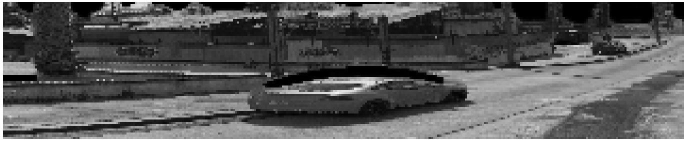
Evaluation

- No sensible metric of “success”
- However, the transformation is probably heading into a direction of the correct solution
- Changing the intensity and depth according to the perceived object

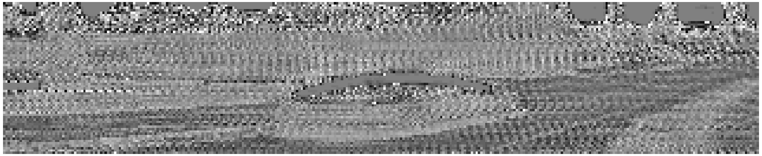
GTA “intensity” (grayscale)



Transformed intensity



Difference image



$$\text{diff} = 10 * (\text{orig} - \text{conv}) + 0.5$$

Conclusion

- We created a pipeline for transforming GTA V data into LiDAR-like data.
- These data were then transformed to be more realistic by CycleGAN.
- The results show that this is a promising and feasible approach even for depth measurements.

Thank you for your attention

Thank you for your attention

Any questions?