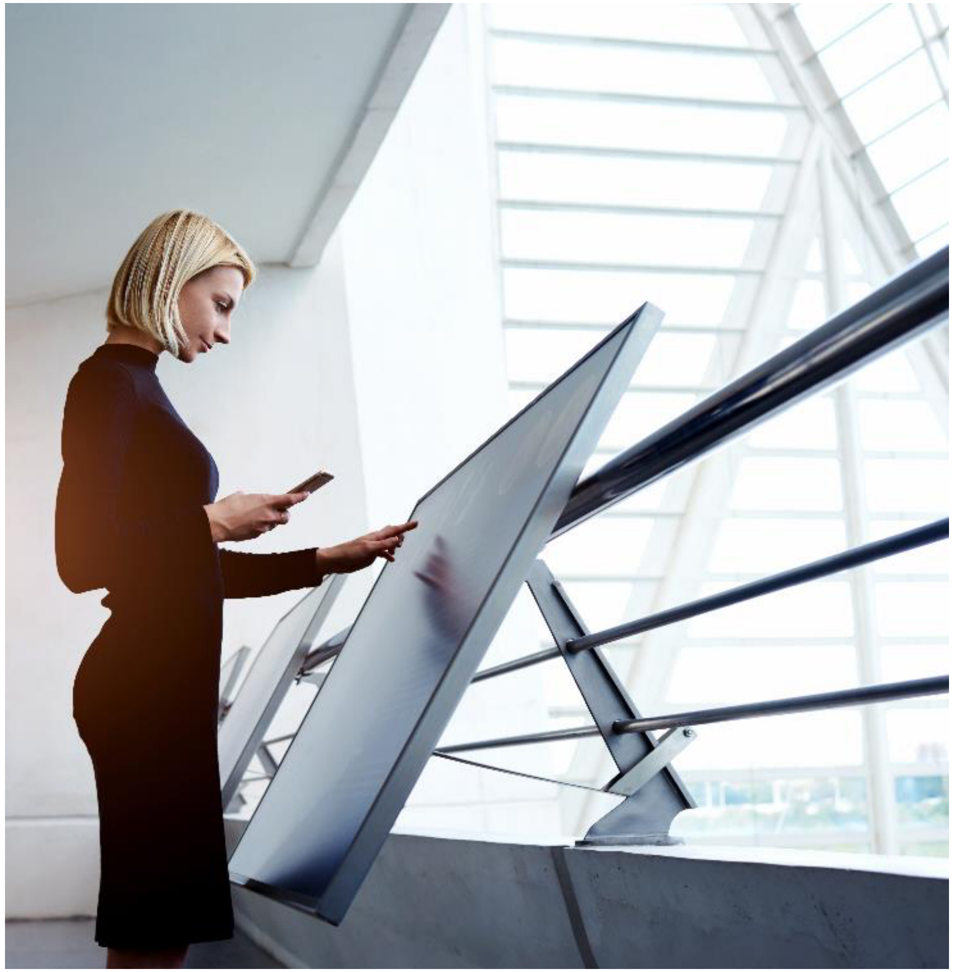
Deloitte.

# DTU Deep Learning Deloitte Projects Kick-off

Segmentation, Synthetic Data & self-driving cars

# Agenda









© 2018 Deloitte Touche Tohmatsu Limited

# Project A: Segmentation of car parts

Project A: Segmentation of car parts

Project B: Synthetic data generation

Project C: Self-driving cars

Groups, projects and next steps



© 2018 Deloitte Touche Tohmatsu Limited

# Why is a segmentation model for car parts useful?

## Each type of data requires different model architectures

**Car part segmentation model:**

Developing a segmentation model for car parts can lead to a number of different downstream tasks – some of which could have high industry impact

**Orientation:**

Having a segmentation model on car parts, it is possible to deduce the orientation of vehicles (which could be useful in various settings involving video)

**Insurance use-case:**

Having a segmentation model on car parts in combination with a model to locate and assess damages could be of huge value in the insurance sector.





© 2018 Deloitte Touche Tohmatsu Limited

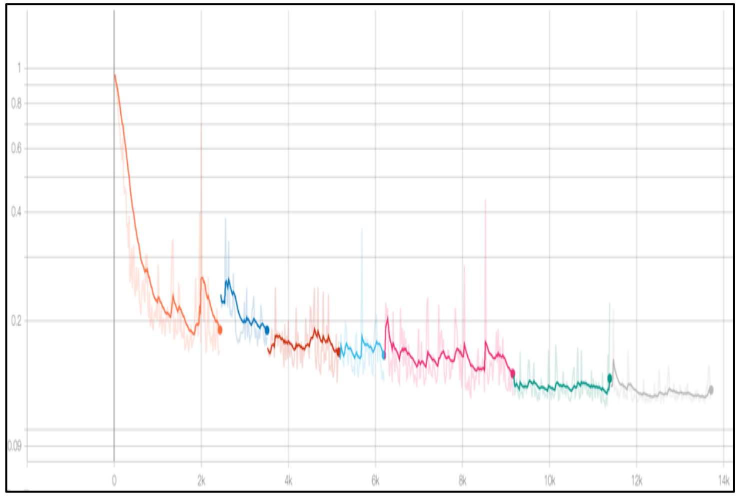
# The potential tasks

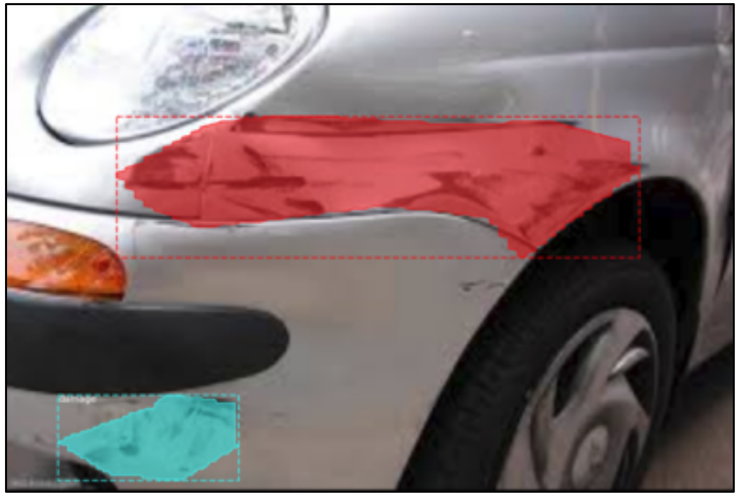
## If time allows it, generating a damage segment

**Working on improving the current Car Segmentation model**

→ Ambitious Groups

**Provided a segmentation model, gather data and train a damage locator**










© 2018 Deloitte Touche Tohmatsu Limited

# The Dataset

## What are you in for and can you improve the dataset even more?

*A Non-Disclosure Agreement has been signed by Ole Winther on your behalf to not share the dataset*

**A record is comprised of an RGB image (.jpg) with dimensions 512 by 256 (2 concatenated 256 by 256 images). The left image represents the target whereas the right image represents the input.**

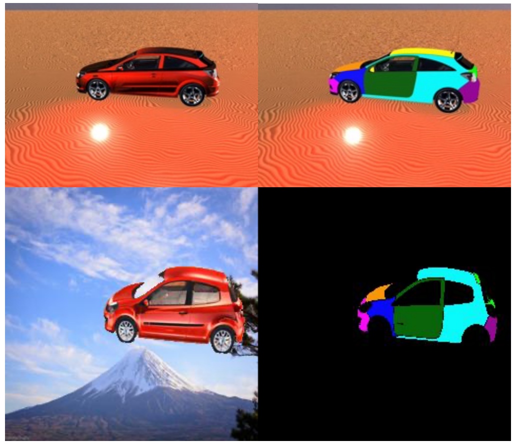
**Hand labelling and CAD combined with Cycle-GAN**

**Original Dataset**

- **1498 distinct datapoints**
- **2918 total datapoints (allowing for scaling, flipping and rotation augmentations)**

+

**Car models rotated in CAD**



**Dataset V2**

- **Random subsample of 2918 cars**
- **Random backgrounds generated using Sobel filters and Watershed segmentation**

Project A: Segmentation of car parts

Project B: Synthetic data generation

Project C: Self-driving cars

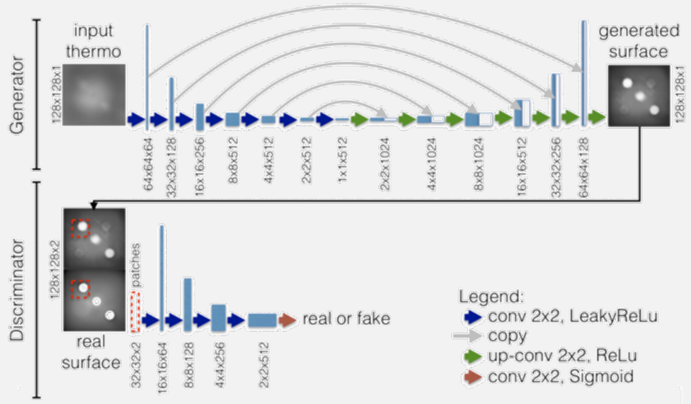
Groups, projects and next steps

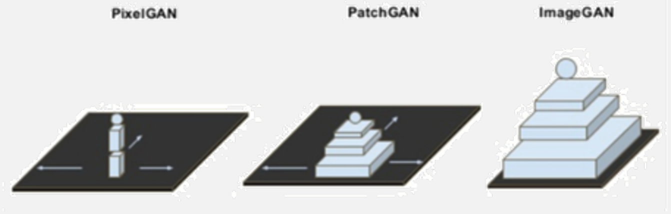
© 2018 Deloitte Touche Tohmatsu Limited

# Inspiration to model architectures

## Ideas for architectures relating to segmentation

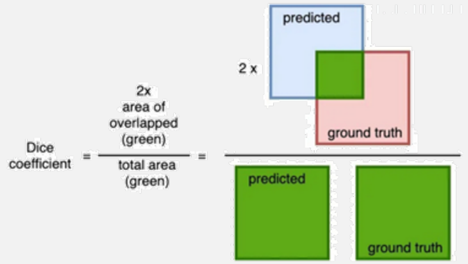
### Pix2Pix GAN:



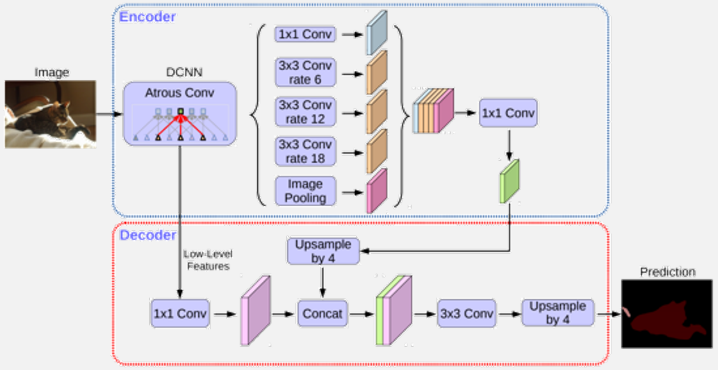




$$G_{Loss} = GAN_{Loss,F} + \gamma \times DICE_{Loss}$$

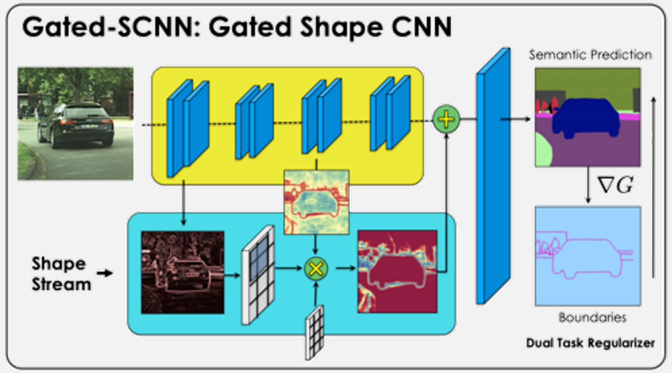
$$D_{Loss} = \frac{(GAN_{Loss,R} + GAN_{Loss,F})}{2}$$




### DeepLabV3+




### Gated-SCNN




Project A: Segmentation of car parts

Project B: Synthetic data generation

Project C: Self-driving cars

Groups, projects and next steps



© 2018 Deloitte Touche Tohmatsu Limited

# Project B: Synthetic data generation for anonymization





© 2018 Deloitte Touche Tohmatsu Limited

# Why is synthetic data generation valuable?

## In Deloitte getting access to our clients data can be tedious and time consuming due to regulatory processes

### Data utility

Synthetic data using generative models have shown enable researchers to **be as efficient** on synthetic data as the original data

### Data privacy

Proper synthetic data generation provides highest guarantees of data privacy enabling easier **data exchange of sensitive data**

### Increase performance

Using synthetic data as data augmentation for other models such as classification models can **increase performance**, especially in **imbalanced datasets**.






© 2018 Deloitte Touche Tohmatsu Limited

# Synthetic data generation project scope

Building a generator model which generates synthetic data which maintains most of data utility as well as high privacy guarantees

| Data Type | Models | Privacy | Utility |
|---|---|---|---|
| **Tabular data**<br>E.g.<br>Kaggle heart disease | **Tabular GANs**<br>link | Outliers, model inference attack, uniqueness | Univariate, bivariate, multivariate, relative prediction error etc. |
| **Time-series Longitudinal**<br>E.g.<br>Traffic data<br>Transactional data, longitudinal patient data | **RNN GANs**<br>link | Outliers, model inference attack, uniqueness | Univariate, bivariate, multivariate, relative prediction error, autocorrelation etc |
| **Text data**<br>E.g.<br>Summary text | **Transformers**<br>link | model inference attack or others? | Sentiment comparisons, relative prediction error, more? |
| **Image segmentation**<br>E.g.<br>car parts segmentation dataset | **Dual GANs**<br>link | model inference attack, Siamese networks? | Relative prediction error, FID, more? |





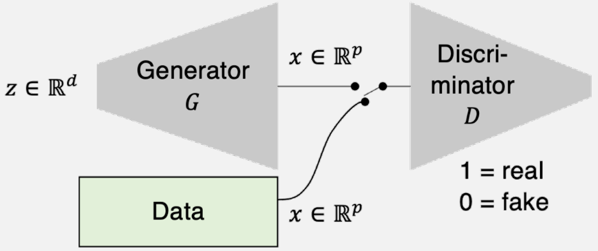
© 2018 Deloitte Touche Tohmatsu Limited

# Inspiration to model architectures

Each type of data requires different model architectures

## Tabular data

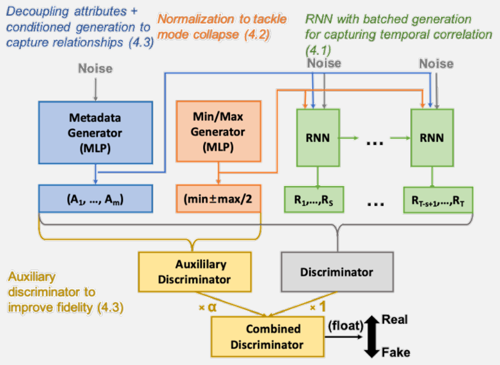
- Discrete attributes using gumbel-softmax
- High dimensionality data




https://arxiv.org/pdf/1909.13403.pdf

## Time series / longitudinal data

- Mode collapses due to high variability
- Long range dependencies




https://arxiv.org/pdf/1909.13403.pdf

## Text data

- Hard to evaluate privacy loss
- Does the text contain all statistical properties (can it be used for sentiment, NER, etc?)
- Transformers are current leading text generators

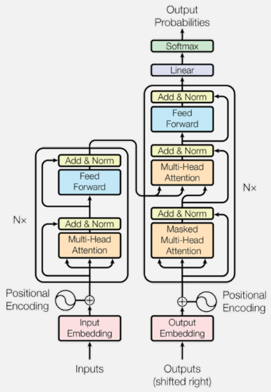
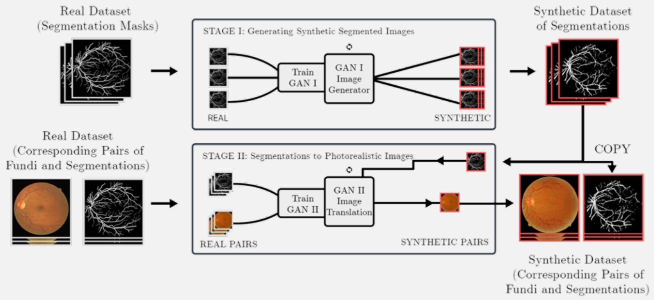



Figure 1: The Transformer - model architecture.

https://arxiv.org/abs/1706.03762

## Image segmentation data

- Complex problem to generate two images
- Simplify the problem




https://arxiv.org/pdf/1709.01872.pdf






© 2018 Deloitte Touche Tohmatsu Limited

# Project C:
# Self-driving cars using Reinforcement Learning or Genetic Algorithms





© 2018 Deloitte Touche Tohmatsu Limited

# The benefit of autonomous vehicles

## The world is becoming increasingly autonomous

*We have several use cases for the self-driving race car. Our general use case is navigation of different vessel types or vehicles: trains can be trained to prevent queues, the flow of insurance claims for large pension funds can be optimized and drones can be trained to search for known objects e.g. surface vessels in the Arctic.*



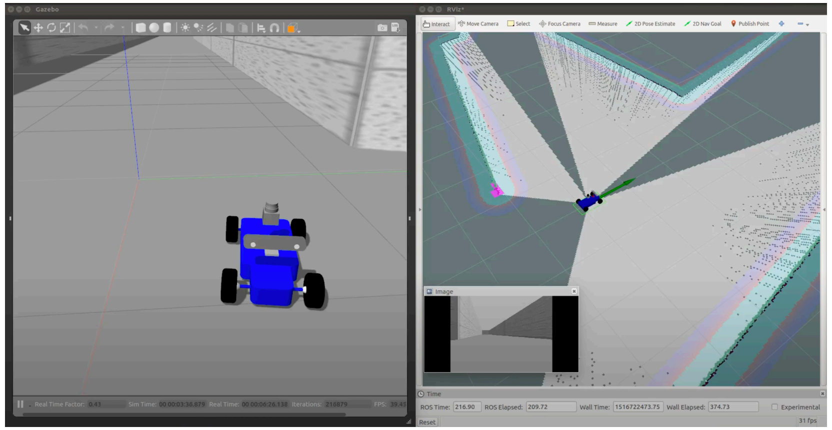


© 2018 Deloitte Touche Tohmatsu Limited

# Environments using for the project

## Using in-house developed simulation code to improve computational time

Full MIT-racecar gazebo environment:



Fast python implementation using basic physics to calculate a new location and lidar feedback based on velocity, acceleration and wheel angle.





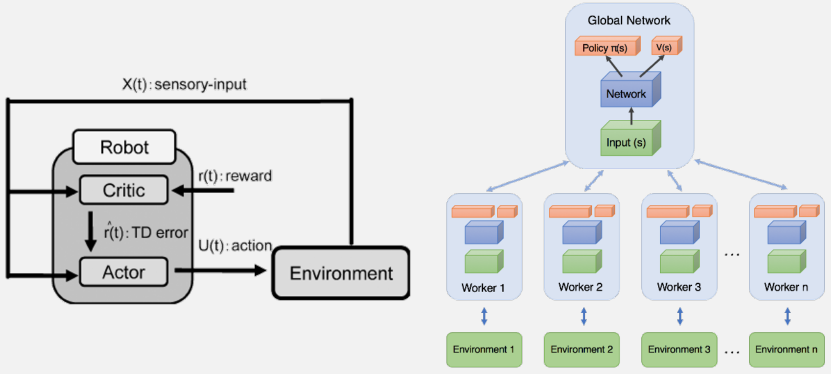
© 2018 Deloitte Touche Tohmatsu Limited

# Possible architectures to use for self-driving char

Using curriculum learning can simplify the task of learning to drive autonomously

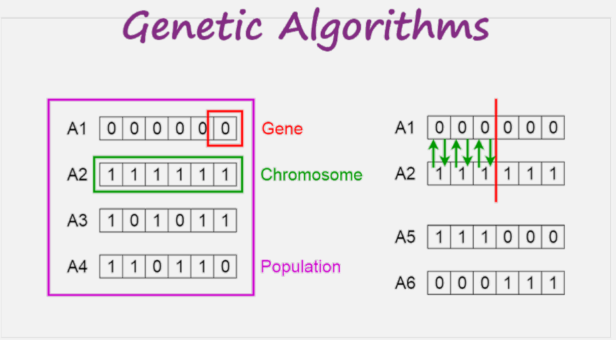
- Only lidar as environmental features (local map information)
- Curriculum learning
  - Drive straight → Avoid single obstacle → navigate maze
- Discrete actions or continuous?
- Genetic algorithms tends learn faster
- Knowledge of previous action
  - Recurrent Neural Networks?

## Deep Reinforcement Learning




https://medium.com/@henrymao/reinforcement-learning-using-asynchronous-advantage-actor-critic-704147f91686

## Genetic Algorithms




https://towardsdatascience.com/introduction-to-genetic-algorithms-including-example-code-e396e98d8bf3





© 2018 Deloitte Touche Tohmatsu Limited

# Groups, projects and next steps

# Groups & Projects

*We would very much prefer to get students on working on all 3 projects as all have great value for us in our business.*

| Group nr. | Teams | Priority | A. Segmentation | B. Synthetic data | C. Race car |
|---|---|---|---|---|---|
| 1 | s151074,s151005,s180014 | A | X | | |
| 2 | s200140, s183307, s175247 | A | X | | |
| 3 | s193204, s193282, s165946 | C | | | X |
| 4 | s153990, s163870, s165206, s153607 | A | X | | |
| 5 | s192142, s192124 | C | | | X |
| 6 | s190328, s192678, s200362 | C | | | X |
| 7 | s193096, s192533, s202451 | C | | | X |
| 8 | s192259, s173922, s184299 | A or C | X | | |
| 9 | s164537. s163740, s174247 | A,B or C | X | | |
| 10 | X | | X | | |





© 2018 Deloitte Touche Tohmatsu Limited

# General information & next steps

## Supervision format

- 1 hour weekly meetings using zoom and break out rooms for each group
  - We will rotate between these groups answering questions
- Communication as well as urgent questions through slack channel & private messages

## Next steps

- All groups write an email to majespersen@deloitte.dk containing the following information and data/simulation code will be distributed according on assigned project
  - Group number
  - Team name (your own to design)
  - Names and student ids
  - Assigned project





© 2018 Deloitte Touche Tohmatsu Limited

# Wrap-up and questions

Thank you for choosing our projects and we are looking forward to follow you all!





© 2018 Deloitte Touche Tohmatsu Limited

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.

Deloitte is a leading global provider of audit and assurance, consulting, financial advisory, risk advisory, tax and related services. Our global network of member firms and related entities in more than 150 countries and territories (collectively, the "Deloitte organization") serves four out of five Fortune Global 500® companies. Learn how Deloitte's approximately 330,000 people make an impact that matters at www.deloitte.com.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms, and their related entities (collectively, the "Deloitte organization"). DTTL (also referred to as "Deloitte Global") and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please seewww.deloitte.com/about to learn more.

© 2020 Deloitte Statsautoriseret Revisionspartnerselskab. For information, contact Deloitte Global.