# Project 1
# Real-time Domain Adaptation in Semantic Segmentation

**Claudia Cuttano - claudia.cuttano@polito.it**
**Gabriele Rosi - gabriele.rosi@polito.it**

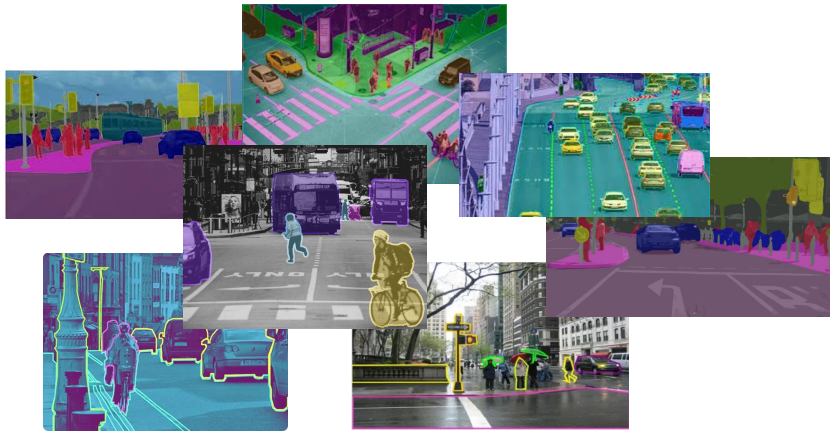

Politecnico di Torino

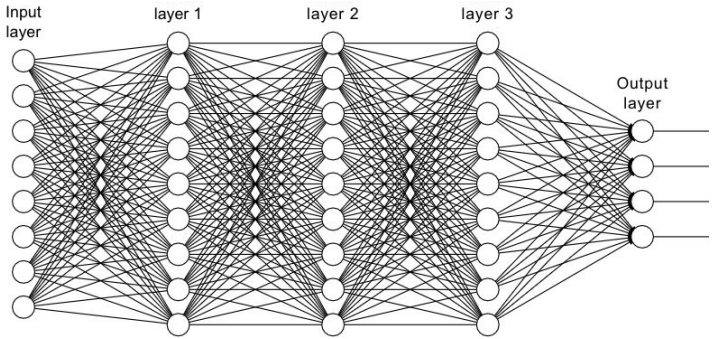
# SEMANTIC SEGMENTATION

- Task of assigning a category label to each pixel of an image.
- Multiple applications: *e.g.* self-driving cars.

# CHALLENGES

- Network performances
- Data labelling





# PERFORMANCE AND EFFICIENCY

- Performance at the expense of efficiency (in terms of parameters, latency and hardware requirements)
  - Large number of parameters
  - High latency
  - Deployments issues on practical scenarios

# PERFORMANCE AND EFFICIENCY

- Performance at the expense of efficiency (in terms of parameters, latency and hardware requirements)
  - Large number of parameters
  - High latency
  - Deployments issues on practical scenarios

→ Solution: use **REAL TIME SEGMENTATION NETWORK**

→ Trade-off between efficiency and complexity

# BURDEN ON LABELING

- Collect pixel-wise annotations is extremely expensive in terms of time.
  - [1] 1.5 h on average to label a single image!

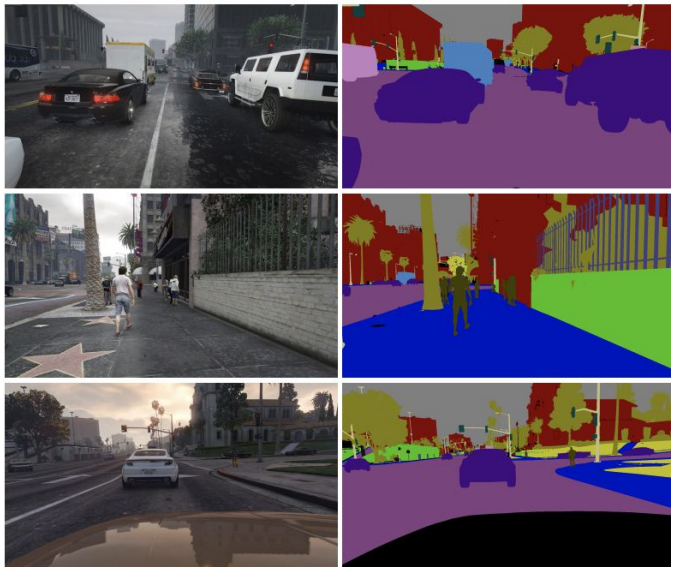
[1] “The cityscapes dataset for semantic urban scene understanding.”, Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele.

# BURDEN ON LABELING

- Collect pixel-wise annotations is extremely expensive in terms of time.
  - [1] 1.5 h on average to label a single image!

→ Solution: use SYNTHETIC DATASETS

- Generated by a software
- Close to real images
- Pixel-wise accurate ground truth
- Cheaper
- *E.g.* images from GTA5 videogame



[1] “The cityscapes dataset for semantic urban scene understanding.”, Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele.

# BURDEN ON LABELING

- Collect pixel-wise annotations is extremely expensive in terms of time.
  - [1] 1.5 h on average to label a single image!

→ **Problem**: gap between artificial images and real images

Source (synthetic) from GTA5

Target (real) image from Turin

# BURDEN ON LABELING

- Collect pixel-wise annotations is extremely expensive in terms of time.
  - [1] 1.5 h on average to label a single image!

→ **Problem**: gap between artificial images and real images

Source (synthetic) from GTA5

Target (real) image from Turin

→ Solution:

**DOMAIN ADAPTATION!**

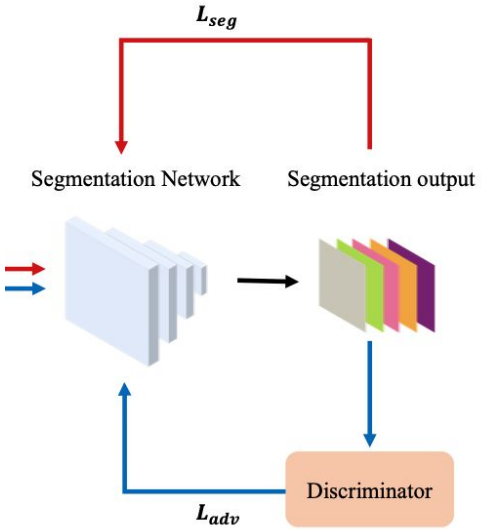
# ADVERSARIAL APPROACH

→ DOMAIN ADAPTATION: **Adversarial Network**

Two components:

- **Segmentation network**
  *Trained to predict the segmentation output for synthetic images.*
- **Discriminator**
  *Distinguish if the prediction derives from a synthetic or real image.*

Objective: fool the discriminator so the prediction for the synthetic and real images are no longer distinguishable.



# IMAGE-TO-IMAGE APPROACH

→ DOMAIN ADAPTATION: **Domain Mixing**


→ DOMAIN ADAPTATION: **Style Transfer**

# YOUR PROJECT

1. **PRELIMINARY ANALYSIS**: Train neural network on the real images (Cityscapes).
    - Standard segmentation network
    - *Real-time* segmentation network

    Evaluate performance, complexity and inference speed.

2. **TRAIN ON SYNTHETIC DATA**
    - Train the real-time network on the synthetic images (GTA5) and evaluate the performance drop when testing this model directly on real ones (Cityscapes).
    - Implement some *data augmentation* techniques when training the network on GTA5

3. **DOMAIN ADAPTATION**
    - Adversarial approach
    - Image-to-image approach

# THANK YOU!

**Claudia Cuttano - claudia.cuttano@polito.it**
**Gabriele Rosi - gabriele.rosi@polito.it**


Politecnico
di Torino