Politecnico di Torino

Approaches for Domain Adaptive Object Detection in Production Environments

Supervisor
Prof. Paolo Garza

Candidate
Gabriele Degola

Company advisors
Neovision
dr. Etienne Balit
ing. Valérian Gonnot

Master of Science in Data Science and Engineering

April 4, 2022

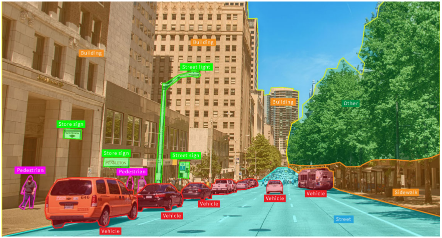
Politecnico di Torino

Contents

Introduction

Introduction

Deep learning methods are extremely effective on computer vision tasks



Introduction

Introduction

- Deep learning methods are extremely effective on computer vision tasks
- ...but data collection and annotation is expensive and time-consuming

Introduction

Introduction

- Deep learning methods are extremely effective on computer vision tasks
- ...but data collection and annotation is expensive and time-consuming
- A solution is to use synthetic data to simulate reality

Introduction

Introduction

- Deep learning methods are extremely effective on computer vision tasks
- ...but data collection and annotation is expensive and time-consuming

- A solution is to use synthetic data to simulate reality
- *Machine learning* models expect training and test data to be drawn from the same data distribution

Introduction

Introduction

- Deep learning methods are extremely effective on computer vision tasks
- ...but data collection and annotation is expensive and time-consuming

- A solution is to use synthetic data to simulate reality
- *Machine learning* models expect training and test data to be drawn from the same data distribution

- **Domain adaptation** methods transfer knowledge between domains

Introduction

Context

- Thesis conducted at Neovision[1], engineering company based in Grenoble, France, and specialized in artificial intelligence

  - Lack of annotated data is a big obstacle for *machine learning* projects
  - Domain adaptation methods depend on specific network architectures, with deep modifications

⇒ Domain adaptation can raise performances, but is complicated to set up in industrial projects

[1] `https://neovision.fr/en/`

Introduction

Method

- Focus on object detection task and *sim2real* transfer
  - Car detection
  - Synthetic dataset: *Sim10k*
  - Real dataset: *CityScapes*
- Research of "simple" approaches for domain adaptation
  - Domain randomization through data augmentation
  - Matching image statistics between domains
- Analysis of popular domain adaptation methods and of their issues
  - One and two stage object detectors: Faster R-CNN[2] and RetinaNet[3]

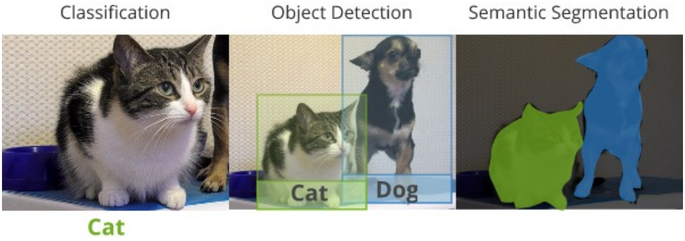
[2] Shaoqing Ren et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2016.

[3] Tsung-Yi Lin et al. *Focal Loss for Dense Object Detection*. 2018.

Domain adaptation

Computer vision

- Convolutional Neural Networks (CNNs) extract features from images
- Three main tasks:
  - Image classification: class of an image
  - Object detection: position and class for all object in an image
  - Semantic segmentation: class of each pixel in an image



Domain adaptation

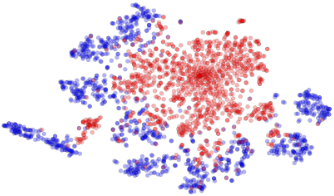
Domain adaptation

- Special case of transfer learning
- Goal to transfer knowledge acquired on a source domain (synthetic images) to a target domain (real images), without training a new model from scratch
- Source domain is annotated, while target domain is not: unsupervised domain adaptation

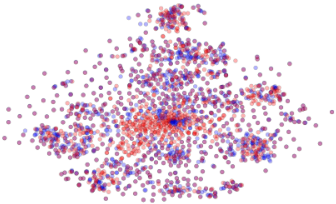
Domain adaptation

Domain adaptation approaches

- ***Domain invariant feature learning***
  - Goal of aligning feature distributions
  - After alignment, it is impossible to determine from which domain features are extracted
  - *Adversarial* approach: feature extractor evolves with the aim of fooling a domain classifier



(a) Non-adapted



(b) Adapted

Domain adaptation

Domain adaptation approaches

- ***Domain invariant feature learning***
  - Goal of aligning feature distributions
  - After alignment, it is impossible to determine from which domain features are extracted
  - *Adversarial* approach: feature extractor evolves with the aim of fooling a domain classifier
- ***Domain mapping***
  - Images from source domain are transformed to look like the target domain
  - Annotations are not altered by the style transfer

Method

Data augmentation for domain randomization

- Data augmentation:
  - Standard technique to ease the training of deep neural networks
  - Training images are randomly modified to introduce new data and increase variability
  - Geometric or color-based

Method

Data augmentation for domain randomization

- Data augmentation:
  - Standard technique to ease the training of deep neural networks
  - Training images are randomly modified to introduce new data and increase variability
  - Geometric or color-based

⇒ **Domain randomization**:
  - Network is "overwhelmed" with widely different versions of the same training images
  - Source data distribution may become closer to the target's
  - No image from target domain is needed

Method

Data augmentation for domain randomization

- Data augmentation:
  - Standard technique to ease the training of deep neural networks
  - Training images are randomly modified to introduce new data and increase variability
  - Geometric or color-based

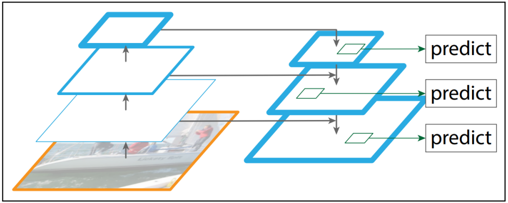
⇒ **Domain randomization**:
  - Network is “overwhelmed” with widely different versions of the same training images
  - Source data distribution may become closer to the target’s
  - No image from target domain is needed

- Experiments confirm the hypothesis:
  - Detection results are improved on *RetinaNet* and *Faster R-CNN*
  - Color operators contribute the most

Method

Adapting the Feature Pyramid Network

- FPN[4] is widely used for object detection:
  - Extract feature maps at multiple levels from a single image
  - Sequence of convolutional layers, size divided by two at consecutive stages, with top-down connections
  - Increases detection performances, but adversarial domain alignment is non-trivial



Method

Adapting the Feature Pyramid Network

- FPN[4] is widely used for object detection:
  - Extract feature maps at multiple levels from a single image
  - Sequence of convolutional layers, size divided by two at consecutive stages, with top-down connections
  - Increases detection performances, but adversarial domain alignment is non-trivial
- Tested approaches:
  - Align levels of extracted feature pyramid
  - Align layers of the ResNet feature extractor, input for FPN
- Extracted representation is meaningful and independent of input domain

[4] Tsung-Yi Lin et al. *Feature Pyramid Networks for Object Detection*. 2017.

Method

Adapting the Feature Pyramid Network

- FPN[4] is widely used for object detection:
  - Extract feature maps at multiple levels from a single image
  - Sequence of convolutional layers, size divided by two at consecutive stages, with top-down connections
  - Increases detection performances, but adversarial domain alignment is non-trivial
- Tested approaches:
  - Align levels of extracted feature pyramid
  - Align layers of the ResNet feature extractor, input for FPN
- Extracted representation is meaningful and independent of input domain

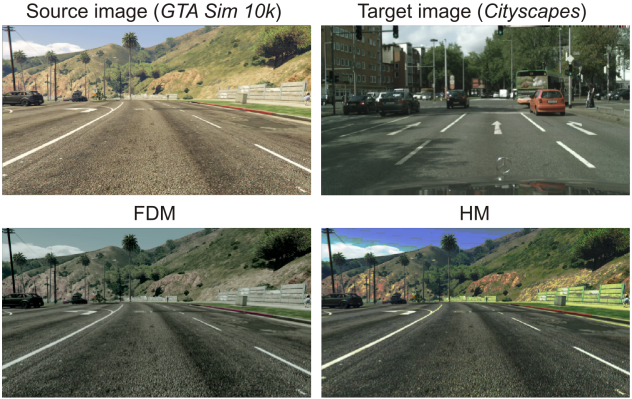
- Method is effective on *RetinaNet*, but not on *Faster R-CNN*

[4] Tsung-Yi Lin et al. *Feature Pyramid Networks for Object Detection*. 2017.

Method

Image statistics matching

- Domain mapping approach that updates source image statistics to match the target domain[5]



[5] Alexey Abramov, Christopher Bayer, and Claudio Heller. *Keep it Simple: Image Statistics Matching for Domain Adaptation*. 2020.

Results

Domain randomization

Transformations						Detector	
Horizontal flip	Rotation	Translation	Color jitter	Solarization	Equalization	mAP@0.5 RetinaNet	mAP@0.5 Faster R-CNN w/ FPN
✓						30.4	46.3
✓	✓	✓				37.6	48.8
✓	✓	✓	✓			49.4	**57.9**
✓	✓	✓	✓	✓	✓	**52.7**	**57.9**
Oracle						68.7	76.3

- FPN drives *Faster R-CNN* to better detection performances than *RetinaNet*

Results

RetinaNet

Method	Transformations				mAP@0.5
	Horizontal flip	Rotation	Translation	Color jitter	
Baseline	✓				30.4
Image statistics matching	✓				48.2
	✓	✓	✓		52.0
	✓	✓	✓	✓	**62.7**
Adversarial adaptation	✓				52.8
	✓	✓	✓		54.2
	✓	✓	✓	✓	54.8
Oracle	✓				68.7

Results

Faster R-CNN

Method	FPN	Transformations: Horizontal flip	Rotation	Translation	Color jitter	mAP@0.5
Baseline		✓				33.7
	✓	✓				46.3
Image statistics matching	✓	✓				55.0
	✓	✓	✓	✓		62.0
	✓	✓	✓	✓	✓	**62.5**
Adversarial adaptation		✓				36.3
	Aligned	✓				45.0
	ResNet aligned	✓				46.8
Oracle	✓	✓				76.3

Conclusions

- Performance gap due to the domain shift between cheap synthetic images and expensive real images

Conclusions

- Performance gap due to the domain shift between cheap synthetic images and expensive real images

- Popular methods for domain adaptive object detection depend on architecture-specific components
- Adversarial alignment may be unstable
- Domain mapping methods are effective with low training and implementation cost: suitable for production environments

Conclusions

- Performance gap due to the domain shift between cheap synthetic images and expensive real images

- 👎 Popular methods for domain adaptive object detection depend on architecture-specific components
- 👎 Adversarial alignment may be unstable
- 👍 Domain mapping methods are effective with low training and implementation cost: suitable for production environments

- Domain randomization through data augmentation can be enough
- ...and can support more advanced domain adaptation methods

Thank you for your attention!

Politecnico di Torino

1859