DOMAIN ADAPTIVE VISUAL OBJECT DETECTION

LEONARDO DI GIOVANNA
GIUSEPPE MOSCARELLI
GAETANO RICCARDO RICOTTA
ANDREA VARA

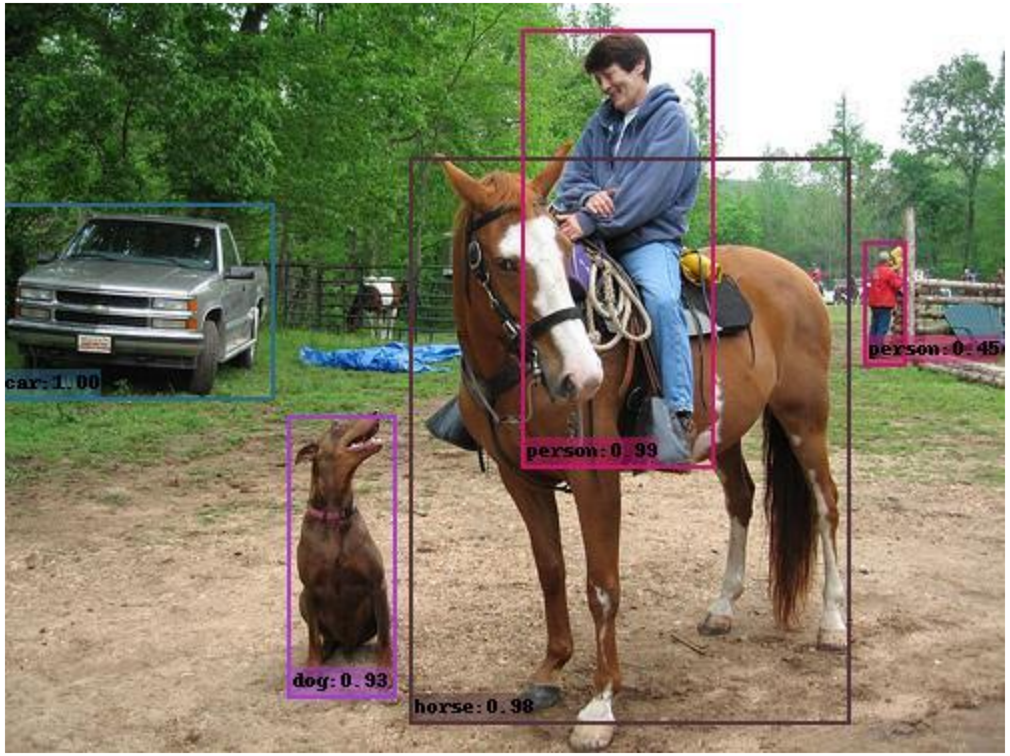
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
2020/21

Politecnico
di Torino
1859

OBJECT DETECTION

Definition

The **Object Detection** task is based on the identification of objects instances belonging to a predefined set of known classes.



OBJECT DETECTION

Detector categories

- Two Stage Object Detectors
- Single Stage Object Detectors

Two stages

- region proposal
- classification and bounding-box regression

Main Solutions

- R-CNN
- Fast R-CNN
- Faster R-CNN

One stage

- localization and content prediction

Main Solutions

- YOLO
- SSD

OBJECT DETECTION

Main problems

- Data amount
- Annotations
- Cross-Domain Generalization

OBJECT DETECTION

A possible solution

Domain Adaptation

Domain Adaptation is a transfer learning subcategory whose goal is to fit a model trained on a specific source domain to a specific target domain.

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

A related work

- *N. Inoue, R. Furuta, T. Yamasaki, and K. Aizawa.* ***Cross-Domain Weakly-Supervised Object Detection through Progressive Domain Adaptation****. arXiv e-prints, pagearXiv:1803.11365, Mar. 2018*

Objective

The main goal is to detect objects as accurately as possible in the target domain by using sufficient instance-level annotations in the source domain and a small number of image-level annotations in the target domain.

Strategy

Starting from a baseline SSD300 model, the authors perform two steps of domain adaptation by finetuning it with the images obtained exploiting two techniques:

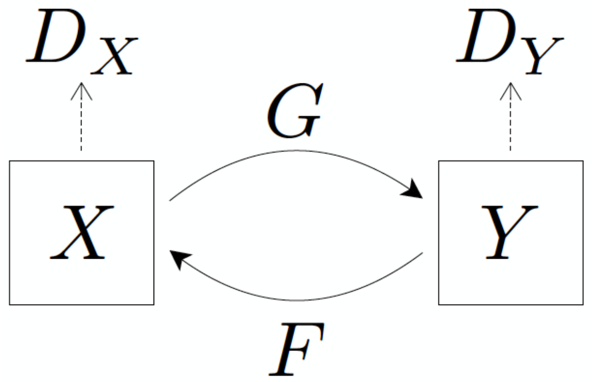
- Domain Transfer (DT)
- Pseudo-Labelling (PL)

Domain Transfer

CycleGAN

$G: X \rightarrow Y$

$F: Y \rightarrow X$



J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. ***Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Net-works****. arXiv e-prints, page arXiv:1703.10593, Mar. 2017*

Pseudo-Labelling

Given the set of all the detections produced by SSD over an image, for each objects class in the image, an annotation corresponding to the top-1 confident detection is assigned.

Source domain

- PASCAL VOC 2007/2012 (20 classes)

Dataset	# images
PASCAL VOC 2007 trainval	5011
PASCAL VOC 2007 test	4952
PASCAL VOC 2012 trainval	11540

Target domain

- Clipart1k
- Watercolor2k
- Comic2k

Dataset	# classes	# images	# instances
Clipart1k	20	1000	3165
Watercolor2k	6	2000	3315
Comic2k	6	2000	6389

Results

Method	SSD300	YOLOv2	Faster R-CNN
Baseline	26.8	25.5	26.2
DT	38.0	31.5	32.1
PL	36.4	34.0	29.8
DT + PL	46.0	39.9	34.9

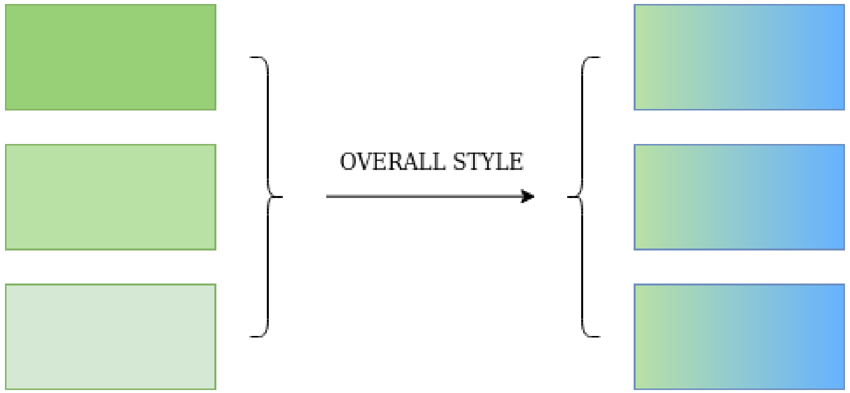
OUR WORK

INTRODUCTION

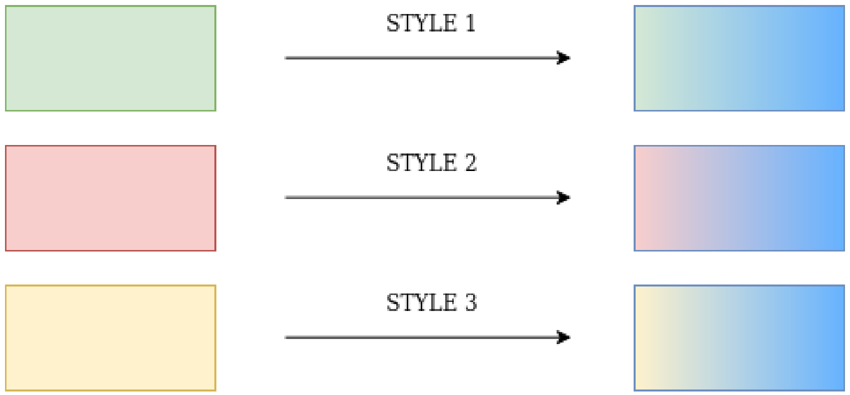
Objectives

- Replicate the results obtained by exploiting the domain transfer technique
- Propose a new adaptation technique based on online style-transfer
- Comparison

Domain Transfer vs. Style Transfer

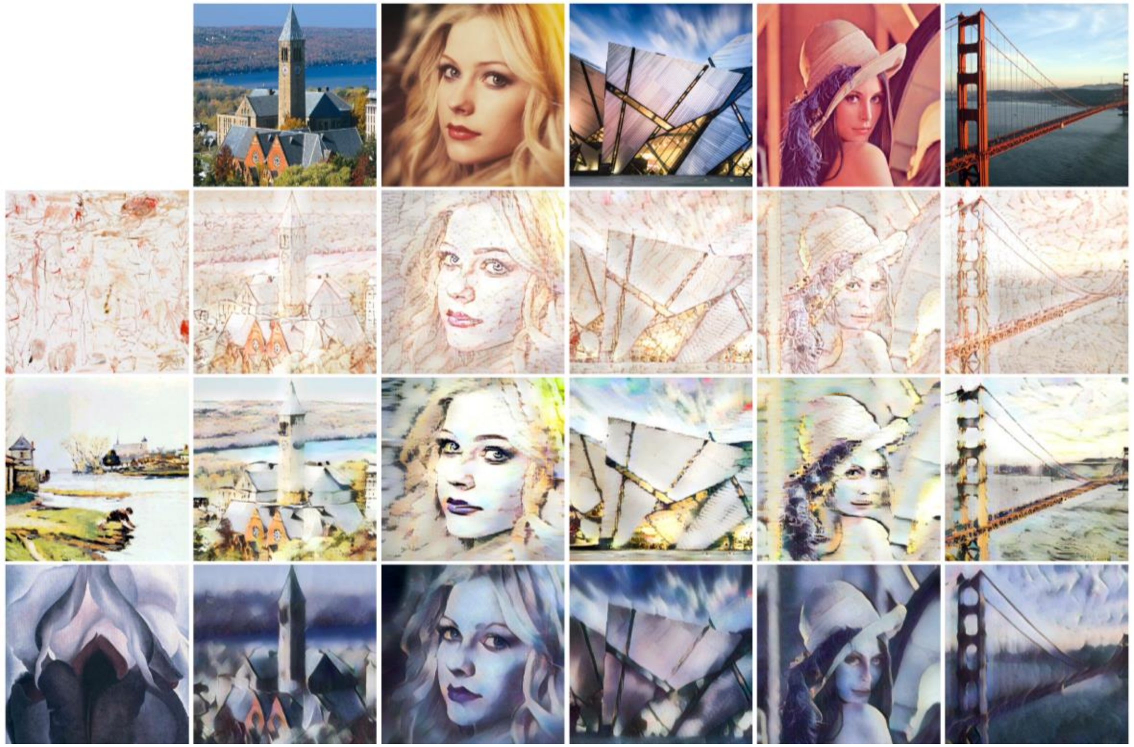


DOMAIN TRANSFER



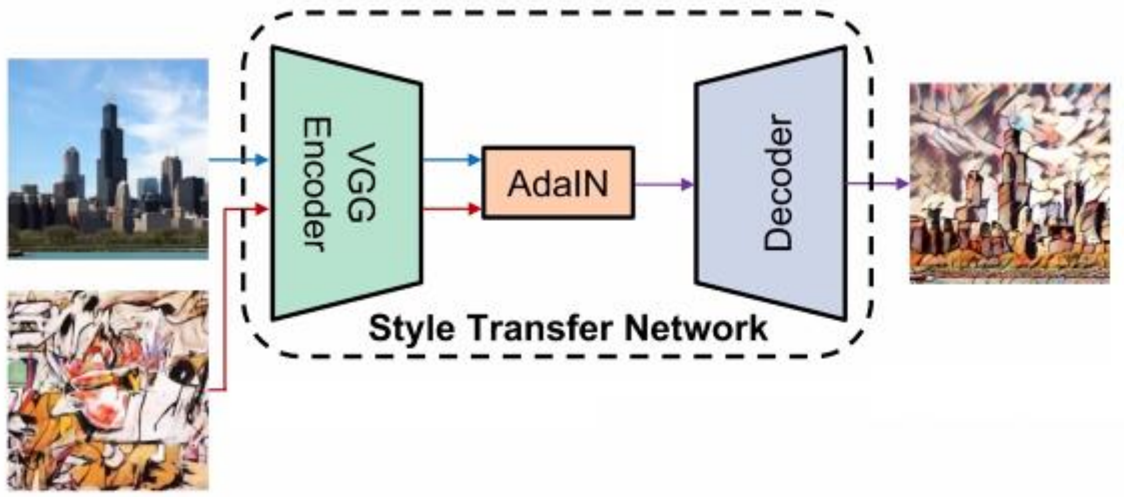
STYLE TRANSFER

Chosen Solution



X. Huang and S. Belongie. ***Arbitrary Style Transfer in Realtime with Adaptive Instance Normalization,*** *arXiv e-prints, page arXiv:1703.06868, Mar. 2017.*

Architecture



X. Huang and S. Belongie. ***Arbitrary Style Transfer in Realtime with Adaptive Instance Normalization****. arXiv e-prints, page arXiv:1703.06868, Mar. 2017.*

AdaIN

AdaIN receives a content input x and a style input y, and simply aligns the channel-wise mean and variance of x to match those of y

$$AdaIN(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

Notation

Content image: c

Style image: s

Encoder Network: $f(\cdot)$

Decoder Network: $g(\cdot)$

AdaIN layer: $AdaIN(\cdot,\cdot)$

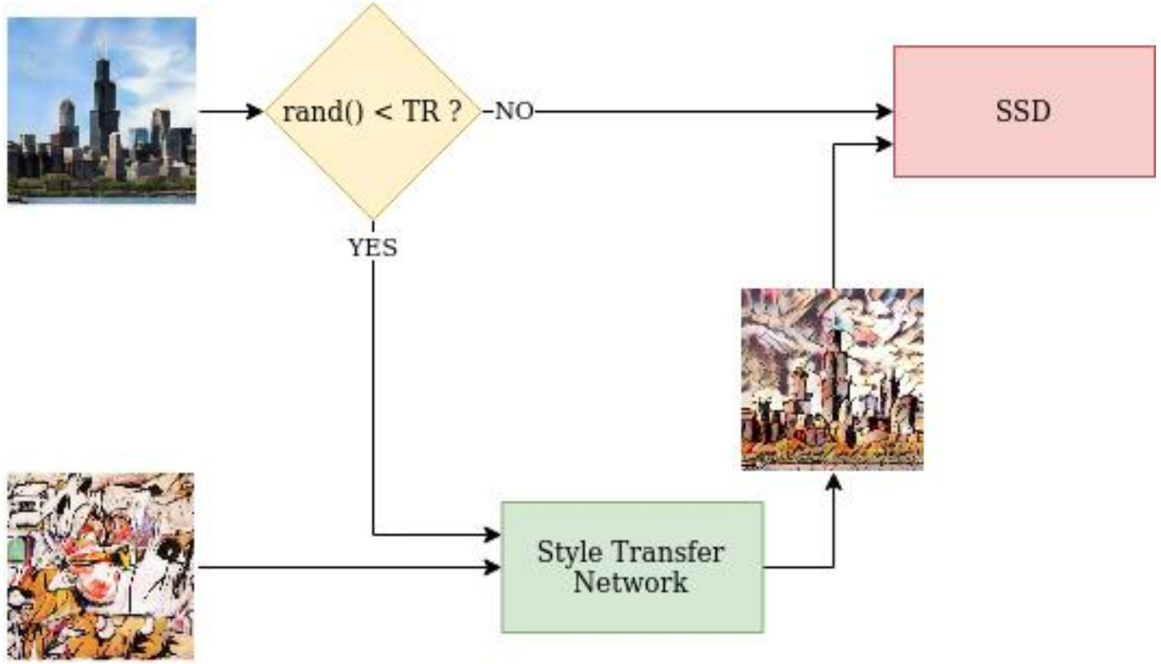
$$T(c,s) = g\left(AdaIN\big(f(c), f(s)\big)\right)$$

Runtime controls: Stylization grade

$$T(c, s, \alpha) = g\left((1 - \alpha)f(c) + \alpha AdaIN\big(f(c), f(s)\big)\right)$$

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

Our work – Domain Adaptation Architecture



Transfer ratio (**TR**) is the probability that a content image is stylized before feeding the SSD

OUR WORK

IMPLEMENTATION DETAILS AND RESULTS

Domains

- **Source domain**
  - PASCAL VOC 2007/12
- **Target domain**
  - Clipart1k

Implementation

- SSD300
  - https://github.com/ekoops/SSD

Datasets

- **Training**
  - PASCAL VOC 2007 trainval split
  - PASCAL VOC 2012 trainval split
- **Test**
  - Clipart1k test split

Details

- Learning rate: 10^{-3}
- Iterations number: 120000
- Momentum: 0.9
- Batch size: 32
- Weight decay: 5×10^{-4}

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

Our work – Baseline SSD model

Results

Model	aereo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mAP
Author's Baseline SSD	19.8	49.5	20.1	23.0	11.3	38.6	34.2	2.5	39.1	21.6	27.3	10.8	32.5	54.1	45.3	31.2	19.0	19.5	19.1	17.9	26.8
Our Baseline SSD	18.9	56.6	19.2	19.9	10.1	26.5	27.0	10.5	42.7	10.3	24.9	14.5	30.9	48.5	35.0	30.4	9.1	24.2	26.5	22.5	25.4

Implementation

- CycleGAN
  - https://github.com/ekoops/pytorch-CycleGAN-and-pix2pix

Datasets

- **TrainA**
  - PASCAL VOC 2007 trainval split
  - PASCAL VOC 2012 trainval split
- **TrainB**
  - Clipart1k train split

Details

- Total epochs: 20
- Learning rate
  - 10^{-5} for the first 10 epochs
  - Linear decay rate to 0 for the last 10 epochs
- λ: 10
- Batch size: 1

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

Our work – Baseline SSD model finetuning with domain-transferred images

Datasets

- **Training**
  - PASCAL VOC 2007 trainval split domain-transferred
  - PASCAL VOC 2012 trainval split domain-transferred
- **Test**
  - Clipart1k test split

Details

- Learning rate: 10^{-5}
- Iterations number
  - 520 (about 1 epoch)
  - 1100
  - 10000
- Batch size: 32

Our work – Baseline SSD model finetuning with domain-transferred images

Results

Model	# Iterations	mAP
Author's FT	1 epoch	38.0
Our FT	520	30.7
Our FT 2	1100	31.6
Our FT 3	10000	31.7

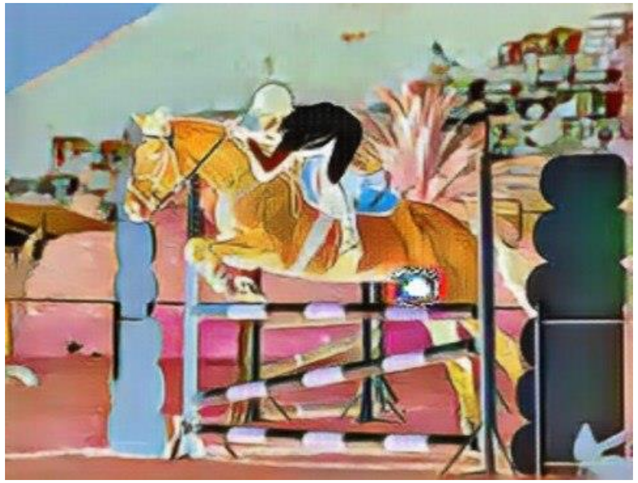
DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

Our work – Baseline SSD model finetuning with domain-transferred images

Possible performance drop reason

Real

Our result



Authors' result

Implementation

- AdaIN
  - https://github.com/ekoops/pytorch-AdaIN

Datasets

- **Training**
  - PASCAL VOC 2007 trainval split
  - PASCAL VOC 2012 trainval split
- **Style**
  - Clipart1k train split
- **Test**
  - Clipart1k test split

Our work – Baseline SSD model finetuning with online style-transferred images

Details

- Images have been resized to 512×512
- Learning rate: 10^{-5}
- Iterations number: $\{520, 1040\}$
- Transfer ratio: $\{0.5, 1.0\}$
- α: $\{0.5, 1.0\}$
- Batch size: 32

DOMAIN ADAPTATION IN VISUAL OBJECT DETECTION

Our work – Baseline SSD model finetuning with online style-transferred images

Results

# Iterations	Transfer ratio	α	mAP
520	0.5	0.5	31.3
520	1.0	0.5	32.7
520	0.5	1.0	36.8
520	1.0	1.0	38.2
1040	1.0	1.0	38.5

Comparison

Model	mAP
Author’s Baseline SSD	26.8
Our Baseline SSD	25.4

Model	mAP
Author’s DT	38.0
Our Proposal	38.2

What could be improved in the future?

- A different Style Transfer Network

THANKS FOR YOUR ATTENTION