Weakly Supervised Object Detection in Artworks

VISART IV Where Computer Vision Meets Art

Nicolas Gonthier

Université Paris Saclay - Télécom ParisTech - LTCI

Joint work with Y. Gousseau, S. Ladjal and O. Bonfait

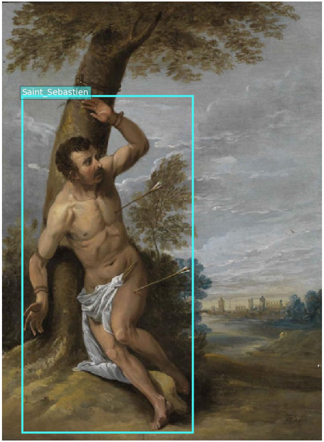
9th September 2018

Motivation

Motivation

Help to search artwork databases.

We would like to **localize** the object of interest



Motivation II

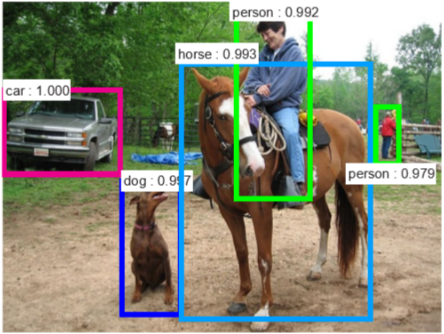
- Use only **image level annotation** → **Weakly supervised** setup
- Fast → No Fine Tuning

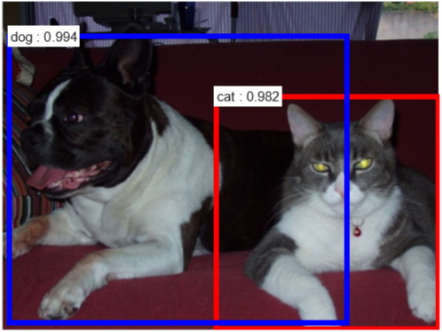
Example images from the IconArt (new) database, for the Saint Sebastian category.

Weakly supervised detection by transfer learning

Implementation details

Use of Faster R-CNN network [Ren et al., 2015] **pre-trained on photography**

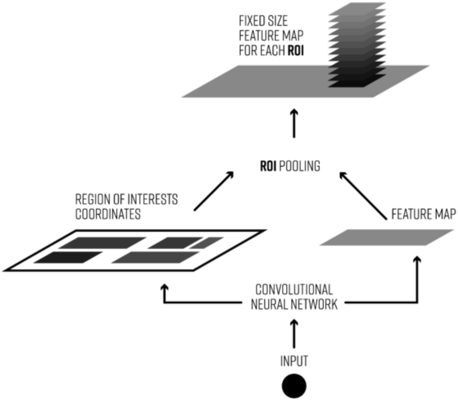




Source: [Ren et al., 2015]

Detection Network

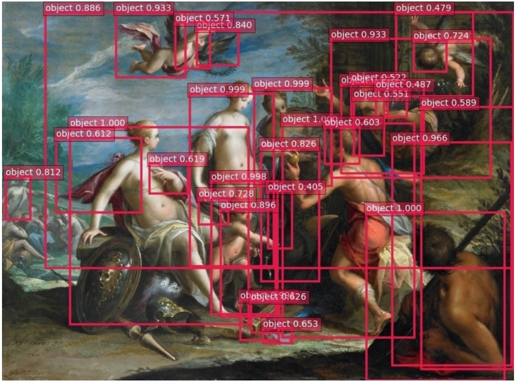
Faster R-CNN network [Ren et al., 2015]

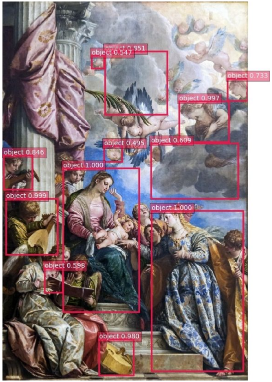


Source: `deepsense.ai`

Multiple Instance Learning Approach

To solve this weakly supervised problem, we use the **Multiple Instance Learning** paradigm. → Regions of an image = bag of elements





Some of the regions of interest generated by the region proposal part (RPN) of Faster R-CNN.

Multiple Instance Learning Approach

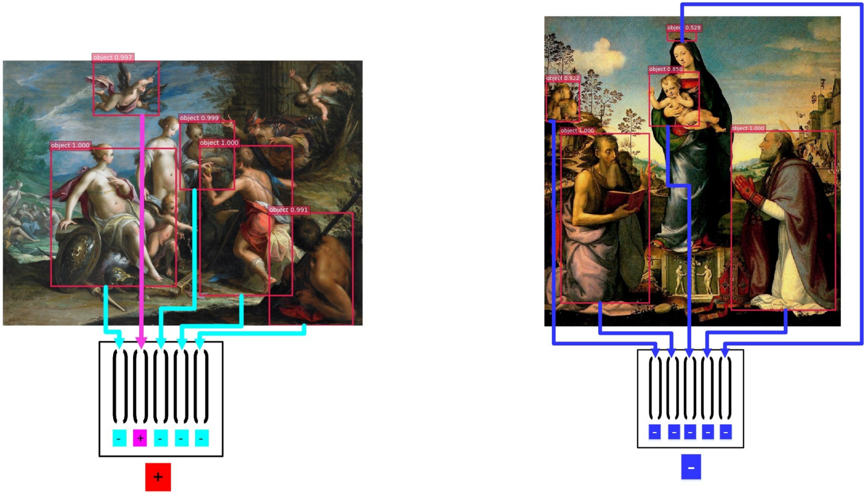
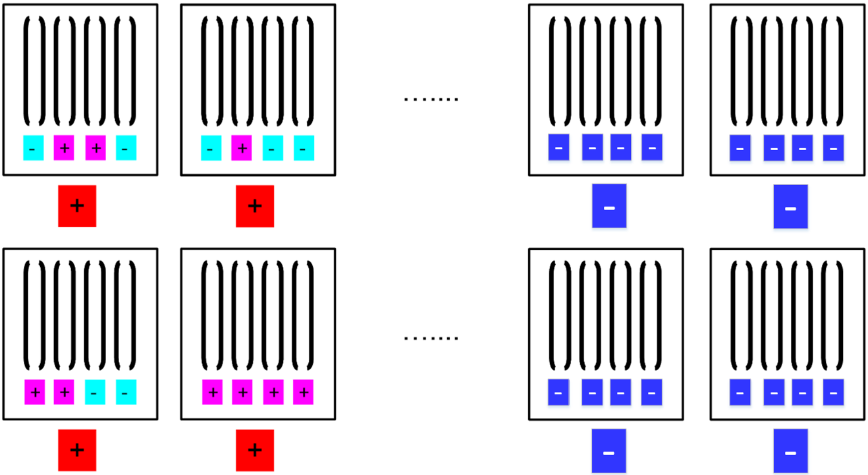


Illustration of positive and negative sets of detections (bounding boxes) for the angel category.

Multiple Instance Learning Approach



How to find the positive vector in each positive bag ?

Source: [Chai, 2011]

How to choose the right region ?

Fine tune Fully Supervised People Detection Network by [Westlake et al., 2016]

DT+PL Cross Domain Weakly Supervised Objects Detection in Watercolor by [Inoue et al., 2018]

MAX Use the **highest score** region to train a SVM to learn new classes [Crowley and Zisserman, 2016]

WSDDN Weakly supervised Deep Detection Network [Bilen and Vedaldi, 2016]

etc

Model : MI-max

For each image i, we have :

$\{X_{i,k}\}_{\{1..K\}}$ features vectors

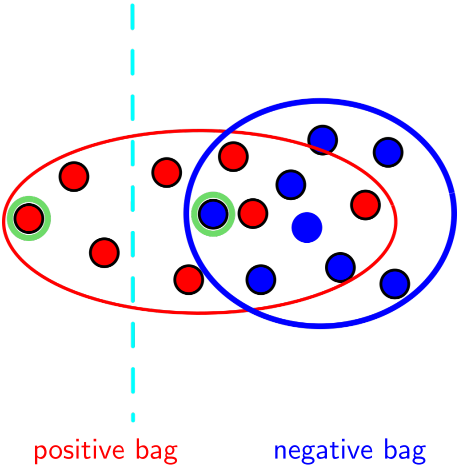
$y_i = \pm 1$ a label

We look for $w \in \mathbb{R}^M$, $b \in \mathbb{R}$ minimizing :

$$\mathcal{L}(w,b) = \underbrace{\sum_{i=1}^{N} \frac{-y_i}{n_{y_i}} Tanh \left\{ \max_{k \in \{1..K\}} \left(w^T X_{i,k} + b \right) \right\}}_{\text{classification loss}} \qquad \underbrace{+ C * ||w||^2}_{\text{regularisation term}} \qquad (1)$$

Simplified version of MI-SVM [Andrews et al., 2003] or Latent SVM [Felzenszwalb et al., 2010].

Model II : MI-max



- positive instance
- negative instance
- Instance used during training step

Model III : MI-max

Use of the objectness score $s_{i,k}$ of each Region of Interest.

$$\mathcal{L}^s(w,b) = \sum_{i=1}^{N} \frac{-y_i}{n_{y_i}} Tanh \left\{ \max_{k \in \{1..K\}} \left(\left(s_{i,k} + \epsilon\right) \left(w^T X_{i,k} + b\right)\right) \right\} + C * \|w\|^2 \quad (2)$$

With $\epsilon \geq 0$.

We do 12 restarts, and select the best couple $(w^\star, b^\star)$.

Test time score for a region x :

$$S(x) = Tanh\{(s(x) + \epsilon)\left({w^\star}^T x + b^\star\right)\} \quad (3)$$

Experiments

Detection evaluation on Watercolor2k [Inoue et al., 2018]

Watercolor2k (test set) *Average precision (%). Comparison of the proposed MI-max method to alternative approaches.*

Method	Net	bike	bird	car	cat	dog	person	mean
DT+PL [1]	SDD	76.5	54.9	46.0	37.4	38.5	72.3	54.3
WSDDN [2]	VGG	1.5	26.0	14.6	0.4	0.5	33.3	12.7
MAX [3]	RES-152-COCO	74.0	34.5	26.8	17.8	21.5	21.0	32.6
MI-SVM [4]	RES-152-COCO	66.8	23.5	6.7	13.0	8.4	14.1	22.1
Our MI-max[5]	RES-152-COCO	85.2 ± 2.5	48.2 ± 1.3	49.2 ± 2.5	31.0 ± 2.5	30.0 ± 2.5	57.0 ± 3.8	50.1 ± 1.1

[1] Method [Inoue et al., 2018].

[2] Method from [Bilen and Vedaldi, 2016] but the performance comes from the paper [Inoue et al., 2018].

[3] [Crowley and Zisserman, 2016]

[4] [Andrews et al., 2003]

[5] Standard deviation computed on 100 runs of the algorithm.

Detection evaluation on People-Art [Westlake et al., 2016]

People-Art (test set) *Average precision (%). Comparison of the proposed MI-max method to alternative approaches.*

Method	Net	person
Fine-tune [6]	Fast R-CNN (VGG16)	59
MAX[7]	RES-152-COCO	25.9
MI-SVM[8]	RES-152-COCO	13.3
Our MI-max	RES-152-COCO	55.4 ± 0.7

[6] The performance comes from the original paper [Westlake et al., 2016].

[7] [Crowley and Zisserman, 2016]

[8] [Andrews et al., 2003]

Watercolor2k and PeopleArt Test examples

Successful examples using our MI-max detection scheme. We only show boxes whose scores are over 0.75.

IconArt : our new database

Class	Angel	Child Jesus	Crucifixion	Mary	nudity	ruins	Saint Sebastian	None	Total
Train	600	755	86	1065	956	234	75	947	2978
Test for detection	261	313	107	446	403	114	82	623	1480
Number of instances	1043	320	109	502	759	194	82		3009

Statistics of the IconArt database

Training examples for the crucifixion and Saint Sebastian categories.

IconArt : Training examples

Example images from the IconArt (new) database, for the angel category.

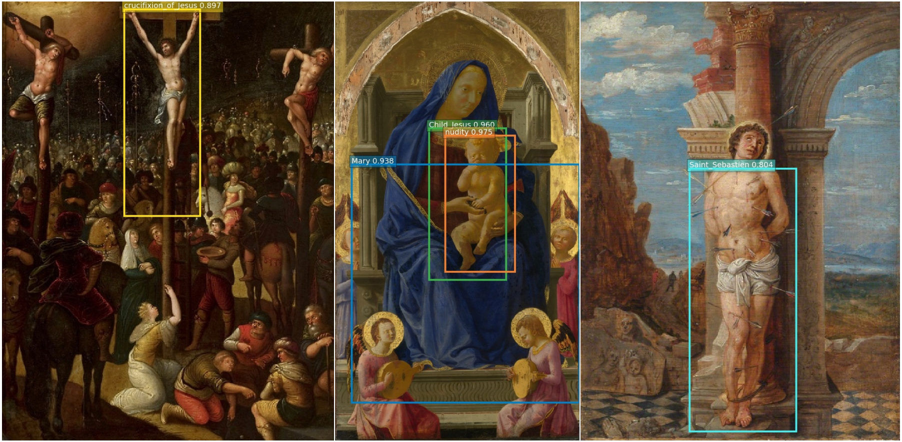
Experiments on IconArt

IconArt detection test set *detection average precision (%). All methods based on RES-152-COCO.*

Method	Metric	angel	JCchild	crucifixion	Mary	nudity	ruins	StSeb	mean
MAX[9]	AP IoU $\geqslant$0.1	10.1	36.2	28.2	18.4	14.0	1.6	2.8	15.9
Our MI-max-C	AP IoU $\geqslant$0.1	12.3 ± 5.4	**41.2** ± 11.3	**74.4** ± 1.6	**46.3** ± 1.7	**31.2** ± 1.9	13.6 ± 4.9	16.1 ± 6.1	33.6 ± 2.2

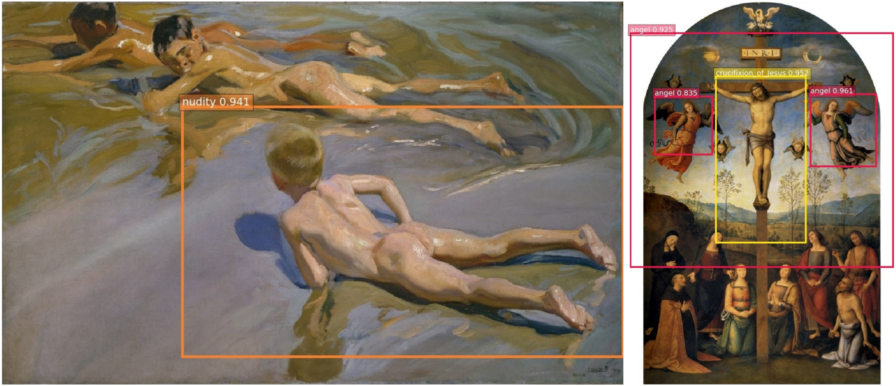
[9] [Crowley and Zisserman, 2016]

Experiments on IconArt, successful examples I



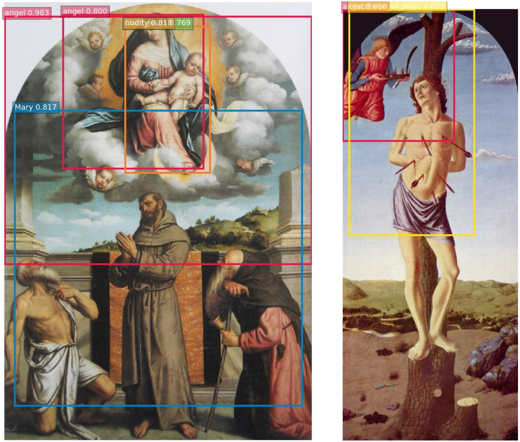
Successful examples using our MI-max-C detection scheme. We only show boxes whose scores are over 0.75.

Experiments on IconArt, successful examples II



Successful examples using our MI-max-C detection scheme.

Experiments on IconArt : failure examples



Failure examples using our MI-max-C detection scheme.

Common Weakly Supervised problems :

- Small discriminative part of the class
- Large portion of the image

Future work

- Promising results on difficult task
- Better understanding of the loss behaviour
- Improve the model
- Fine tune the network
- Use an other pre-trained detection network than Faster R-CNN
- Work on larger databases [Rijksmuseum, 2018, MET, 2018, Réunion des Musées Nationaux-Grand Palais, 2018]

Questions ?

References I

- [Andrews et al., 2003] Andrews, S., Tsochantaridis, I., and Hofmann, T. (2003).
  Support vector machines for multiple-instance learning.
  In *Advances in Neural Information Processing Systems*, pages 577–584.
- [Bilen and Vedaldi, 2016] Bilen, H. and Vedaldi, A. (2016).
  Weakly supervised deep detection networks.
  In *IEEE Conference on Computer Vision and Pattern Recognition*.
- [Chai, 2011] Chai, Y. (2011).
  Support Vector Machines for Multiple-Instance Learning.
- [Crowley and Zisserman, 2016] Crowley, E. J. and Zisserman, A. (2016).
  The Art of Detection.
  In *European Conference on Computer Vision*, pages 721–737. Springer.
- [Felzenszwalb et al., 2010] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010).
  Object detection with discriminatively trained part-based models.
  IEEE transactions on pattern analysis and machine intelligence, 32(9):1627–1645.

References II

- [Inoue et al., 2018] Inoue, N., Furuta, R., Yamasaki, T., and Aizawa, K. (2018).
  Cross-Domain Weakly-Supervised Object Detection through Progressive Domain Adaptation.
  In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*. IEEE.
- [MET, 2018] MET (2018).
  Image and Data Resources | The Metropolitan Museum of Art.
  https://www.metmuseum.org/about-the-met/policies-and-documents/image-resources.
- [Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015).
  Faster r-cnn: Towards real-time object detection with region proposal networks.
  In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc.
- [Réunion des Musées Nationaux-Grand Palais, 2018] Réunion des Musées Nationaux-Grand Palais (2018).
  Images d'Art.
  https://art.rmngp.fr/en.
- [Rijksmuseum, 2018] Rijksmuseum (2018).
  Online Collection Catalogue - Research.
  https://www.rijksmuseum.nl/en/research/online-collection-catalogue.

References III

- [Westlake et al., 2016] Westlake, N., Cai, H., and Hall, P. (2016).
  Detecting people in artwork with cnns.
  In *ECCV Workshops*.