The MET Dataset

Instance-Level Recognition Workshop
ICCV' 21

Nikolaos-Antonios Ypsilantis
CTU in Prague

Noa Garcia
Osaka University

Guangxing Han
Columbia University

Sarah Ibrahimi
University of Amsterdam

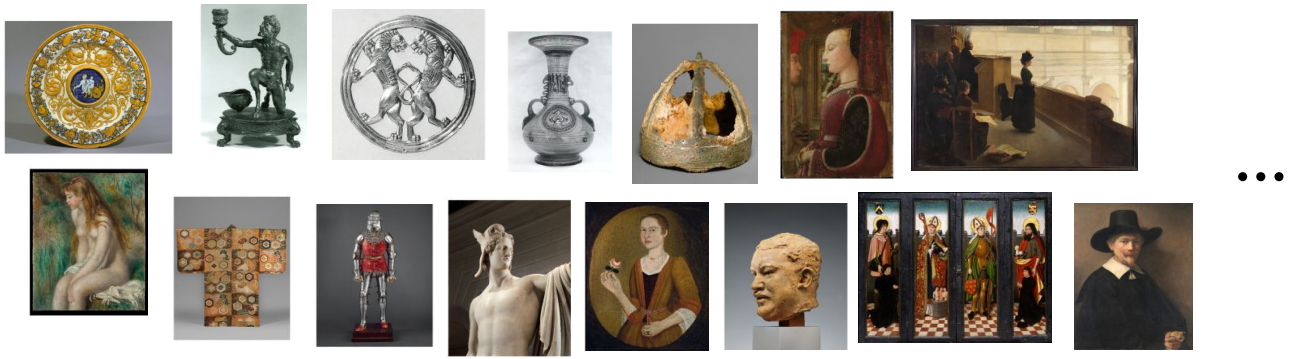
Nanne van Noord
University of Amsterdam

Giorgos Tolias
CTU in Prague

Artwork recognition dataset - motivation

- Real-world applications of artwork recognition
- New domain compared to existing instance-level recognition (ILR) datasets
- Challenging for existing methods

Artwork recognition



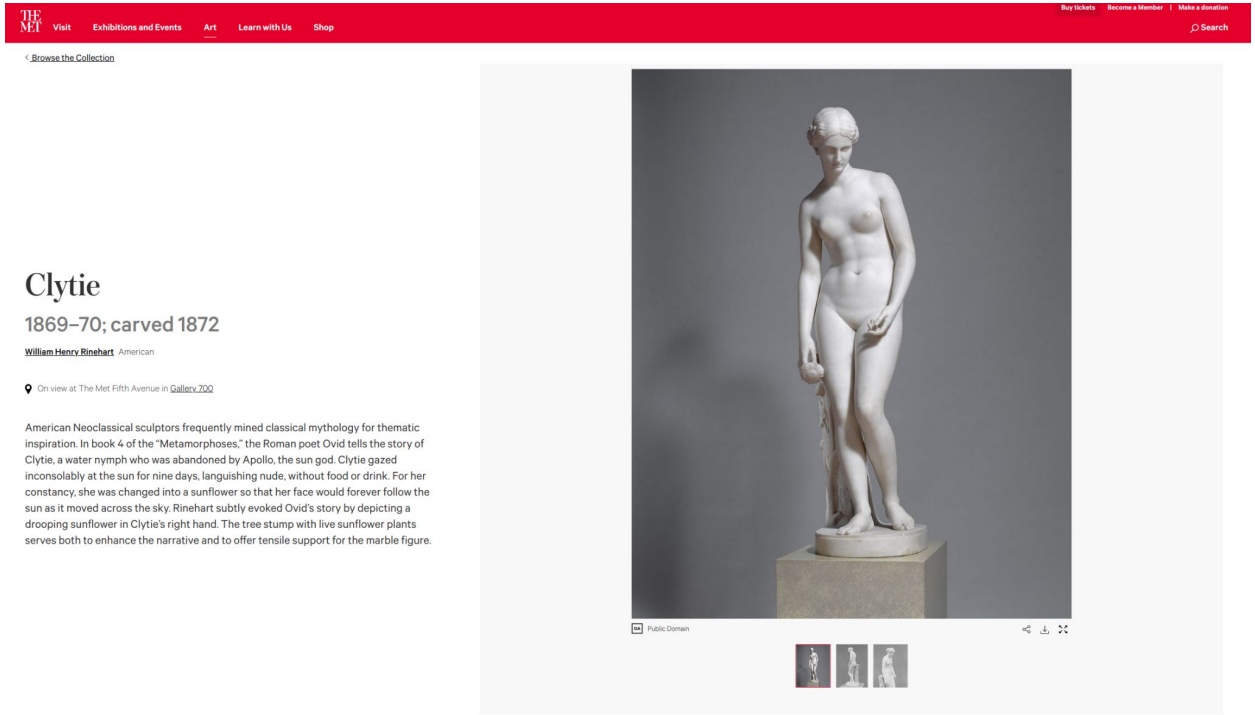
Training set

Query / test

Artwork Classifier

Artwork: California, Hiram Powers, 1850–55
Confidence: 92 %

Metropolitan Museum of New York (MET) - training set

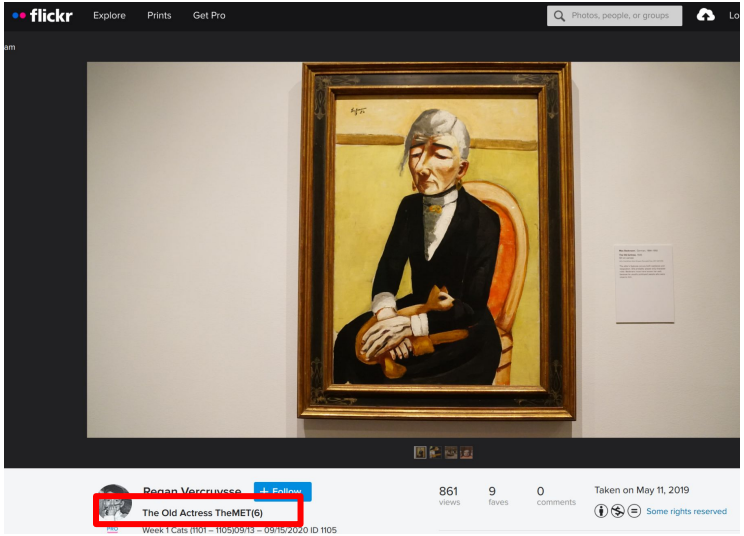


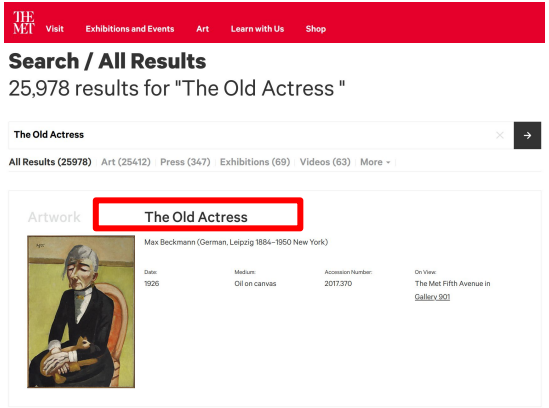
- **Public domain** images (CC0) from The MET open collection
  - Studio conditions
  - Multiple views
- Collection
  - Download all images
  - Automatic deduplication
  - Max 10 images/artwork
  - No extreme aspect ratios
  - Down-sample
- Total
  - ~397k images
  - ~224k artworks (classes)

Queries from The MET

Flickr photos

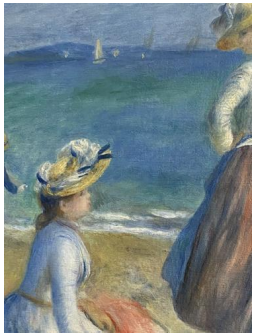
- Crawl MET related photos
- Creative Commons license
- Photos by 38 MET visitors
- Annotation:
  - Text matching
  - Manual filtering





Own photos (Guangxing)

- Faster annotation



Distractor queries

- Follows the paradigm of GLDv2
- Real-world setting:
  - Photos not from “The MET”
- Test robustness to Out of Distribution (OoD) input
- Use Wikimedia commons Public Domain images

Artwork **distractors**

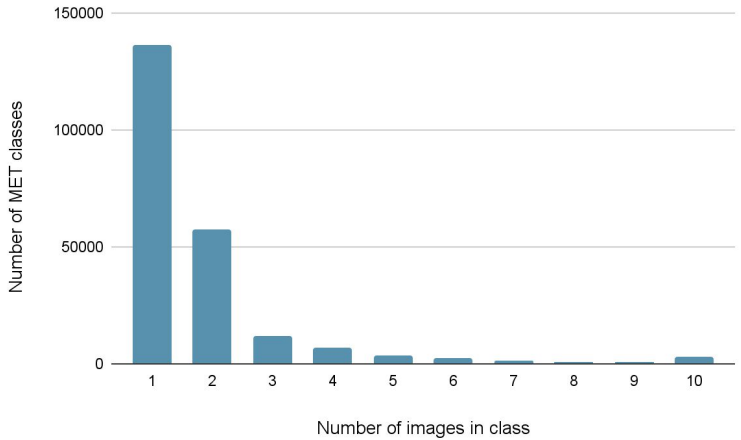
- manual verification

Non - Artwork **distractors**

Dataset statistics - challenges

- Exhibits (train set)
  - 397k images
  - 224k classes
- Queries
  - 21k images
    - 1,1k MET queries
      - 845 classes
      - 14 departments (balanced)
    - 20k distractors
      - Art : 11k
      - No-art : 9k

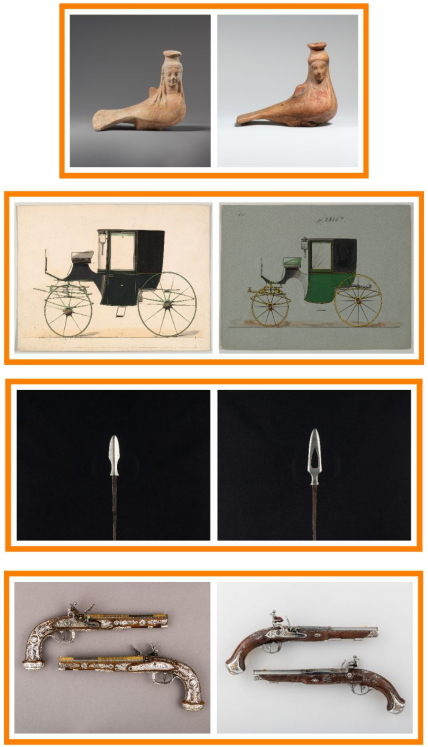
- Large-scale recognition
- Long-tail distribution

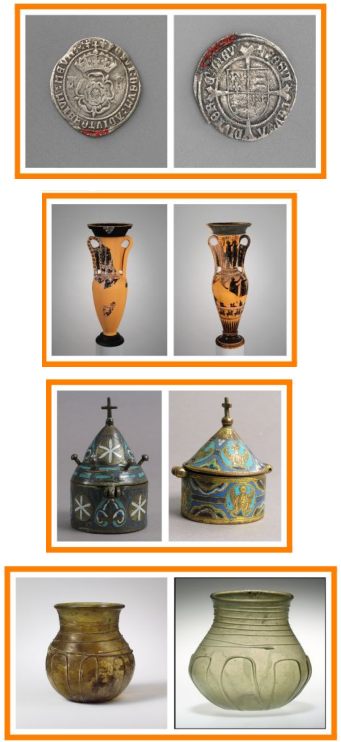


- Distractor (OoD) queries
- High inter-class similarity
- Domain shift between exhibits and queries

High inter-class similarity

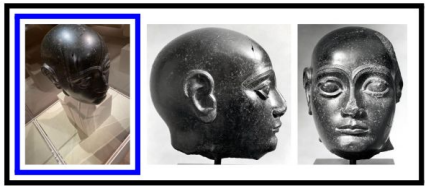
- Examples of visually similar classes





Domain shift

- Viewpoint and illumination changes
- Clutter



Comparison to other datasets

Current Art datasets :

- Attribute prediction
- Category level recognition
- ILR ones:
  - Smaller and noisy

Art datasets	Year	Domain	# Images	# Classes	Type of annotations	Task	Image source
PrintArt [5]	2012	Prints	988	75	Art theme	CLR	Artstor
VGG Paintings [10]	2014	Paintings	8,629	10	Object category	CLR	Art UK
WikiPaintings [20]	2014	Paintings	85,000	25	Style	AP	WikiArt
Rijksmuseum [27]	2014	Artwork	112,039	†6,629	Art attributes	AP	Rijksmuseum
BAM [39]	2017	Digital art	65M	†9	Media, content, emotion	AP, CLR	Enhance
Art500k [26]	2017	Artwork	554,198	†1,000	Art attributes	AP	*Various*
SemArt [14]	2018	Paintings	21,383	21,383	Art attributes, descriptions	Text-image	Web Gallery of Art
OmniArt [35]	2018	Artwork	1,348,017	†100,433	Art attributes	AP	*Various*
Open MIC [22]	2018	Artwork	16,156	866	Instance	ILR (DA)	*Authors*
iMET [41]	2019	Artwork	155,531	1,103	Concepts	CLR	The MET
NoisyArt [11]	2019	Artwork	89,095	3,120	Instance (noisy)	ILR	*Various*
The MET (Ours)	2021	Artwork	418,605	224,408	Instance	ILR	*Various*

Comparison to other datasets

Current ILR datasets :

- Not art
- Large but noisy
- Not fully publicly available

ILR datasets	Year	Domain	# Images	# Classes	Type of annotations	Image source
Street2Shop [17]	2015	Clothes	425,040	204,795	Category, instance	*Various*
DeepFashion [25]	2016	Clothes	800, 000	33,881	Attributes, landmarks, instance	*Various*
GLD v2 [38]	2019	Landmarks	4.98M	200,000	Instance (noisy)	Wikimedia
AliProducts [8]	2020	Products	3M	50,030	Instance (noisy)	Alibaba
Products-10K [2]	2020	Products	150,000	10,000	Category, instance	JD.com
The MET (Ours)	2021	Artwork	418,605	224,408	Instance	*Various*

Benchmark

- Training set: MET open collection images
- Queries are split into test and val sets
  - 90 - 10 % image splits
  - No overlap of photographers or classes
- Test
  - 1,003 MET
    - 25 photographers
  - 18,316 distractors
- Val
  - 129 MET
    - 14 photographers
  - 2,036 distractors
  - Hyperparameter tuning

Evaluation metrics: accuracy and GAP





Accuracy = 66.6%

Sort





Global Average Precision =
⅓ * (1+0.5) = 0.5

confidence:	0.99	0.95	0.85	0.60	0.45
precision:	1	0.5	0.33	0.5	0.4
correct:	1	0	0	1	0





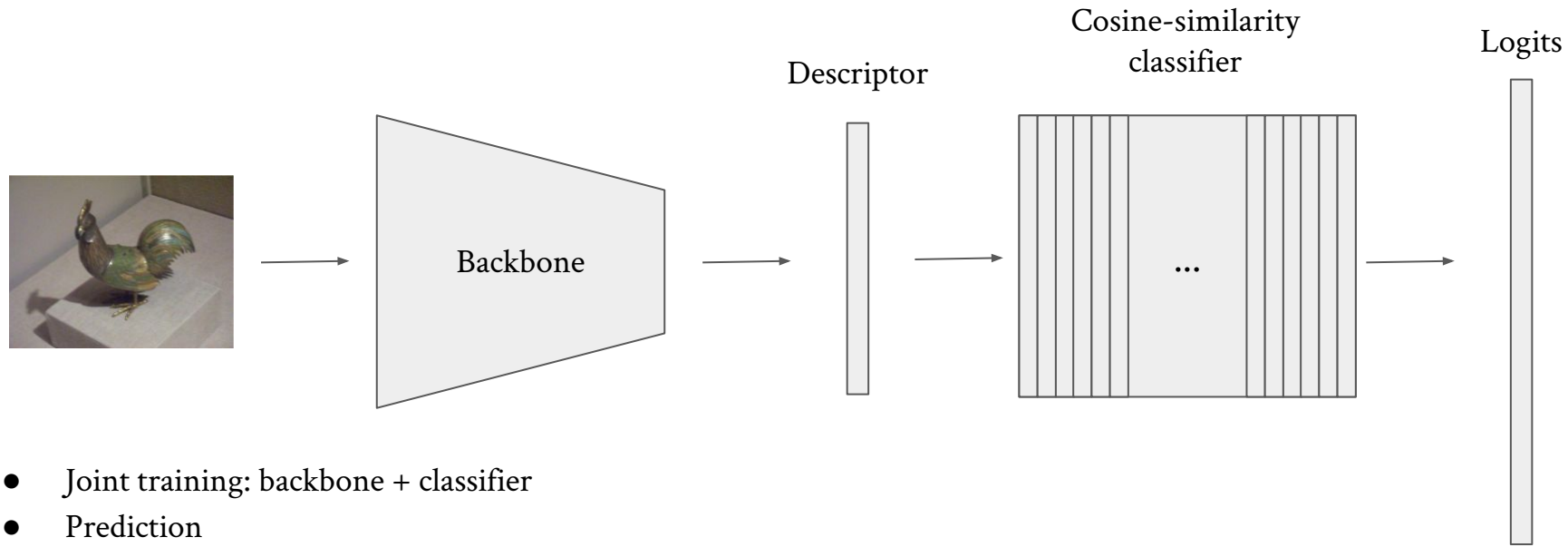
Global Average Precision = 0.81





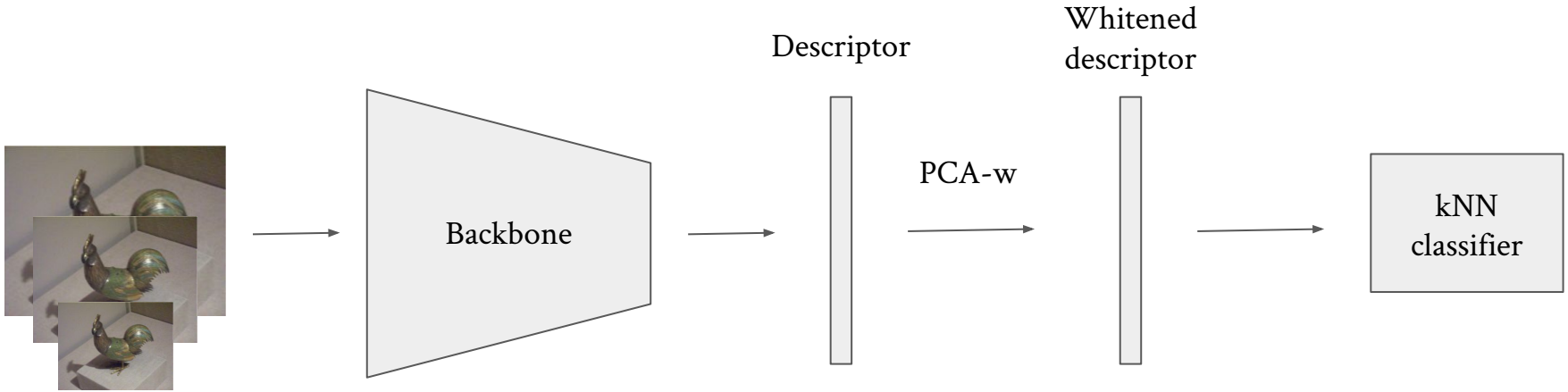
Global Average Precision = 1

Baselines : parametric classification



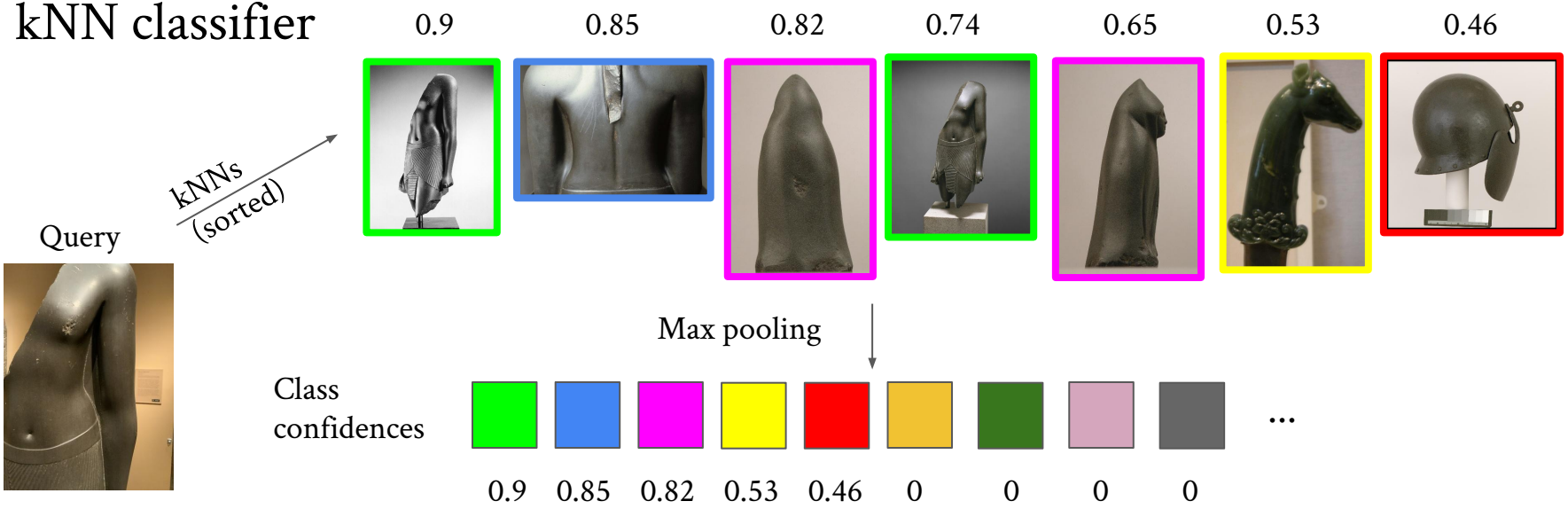
- Joint training: backbone + classifier
- Prediction
  - Argmax
- Prediction confidence
  - Softmax with temperature (τ)

Baselines : non-parametric classification



- Backbone: pre-trained or fine-tuned on MET
- PCA-whitening
  - Learned on train set
- Multiscale representation
  - By sum-aggregation

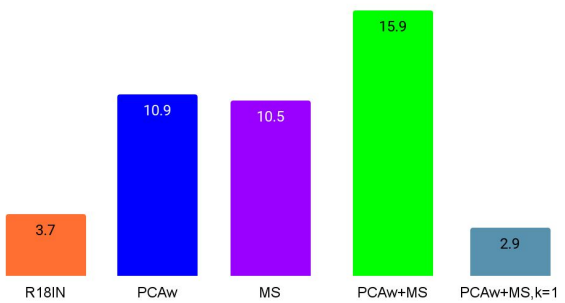
kNN classifier



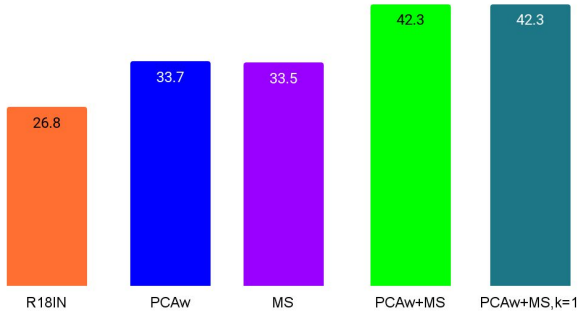
- Prediction
  - First neighbor
- Prediction confidence
  - Normalization by softmax with temperature (τ)
  - Affects GAP, not accuracy
- Number of neighbors (k) and temperature (τ) tuned on val set

Performance evaluation: kNN classifier

GAP



ACC



R18IN:
pre-trained
on ImageNet

no (k,τ) tuning

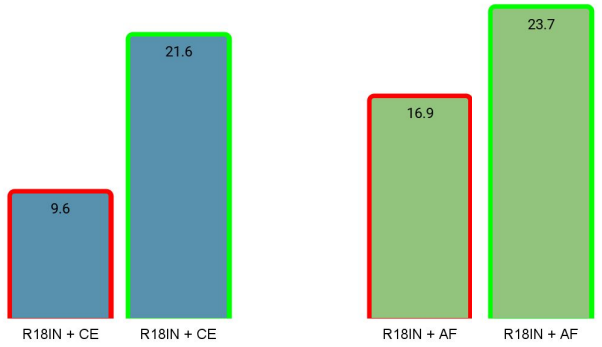
Different pretrained models: kNN classifier

		GAP	ACC
He et al. : Deep residual learning for image recognition.	R50IN	22.2	46.4
Radenović et al. : Fine-tuning cnn image retrieval with no human annotation.	R50SfM	26.6	48.6
Garcia et al. : Context-aware embeddings for automatic art analysis.	R50SemArt (author)	1.8	18.0
	R50SemArt (type)	7.9	31.9
Geirhos et al. : Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.	R50SIN	15.5	41.7
Caron et al. : Unsupervised learning of visual features by contrasting cluster assignments.	R50SwAV	22.8	49.6
Yalniz et al. : Billion-scale semi-supervised learning for image classification.	R50SWSL	**30.4**	**56.3**

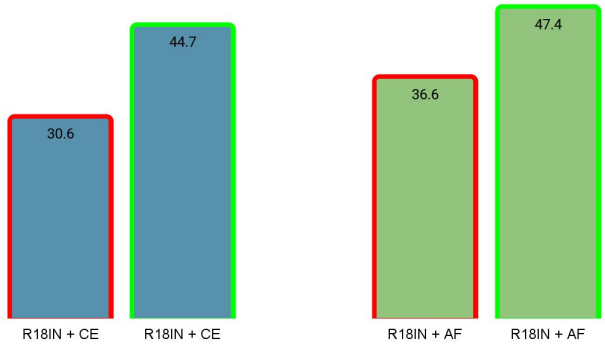
Parametric vs non-parametric classifier

- classification training (parametric)
  - Cross-Entropy
  - Arc-face [Deng et al.]
- use embedding for kNN classification (non parametric)

GAP



ACC



[Deng et al. : Arcface: Additive angular margin loss for deep face recognition]

Representation learning on MET

- SimSiam [Chen et al.]
  - Synthetic positives only

SimSiam method

- Contrastive
  - Hard-negative mining

hard negative

anchor

positive

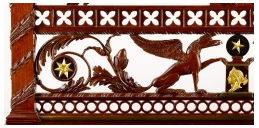
Contrastive loss

candidate positives

case 1 (Con1)

case 3 (Con3)

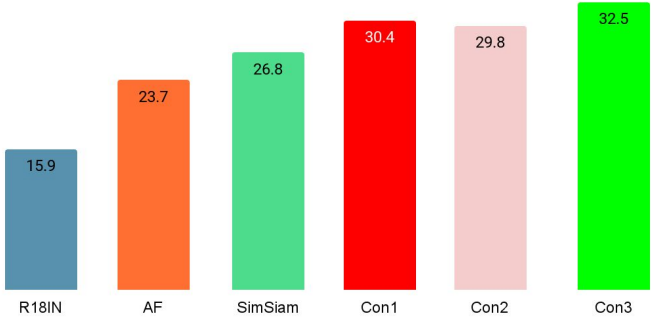
case 2 (Con2)



[Chen et al. : Exploring simple siamese representation learning]

Performance evaluation: representation learning on MET - kNN classifier

GAP

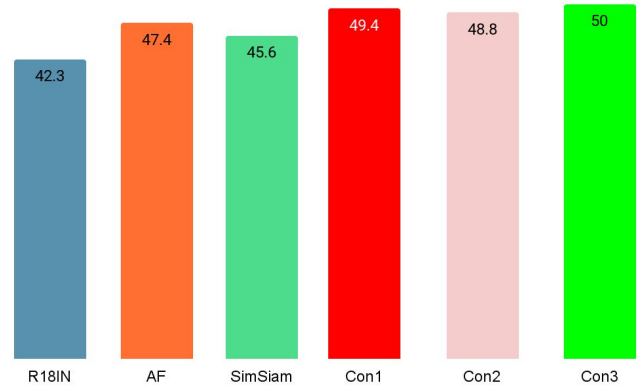


synthetic positives

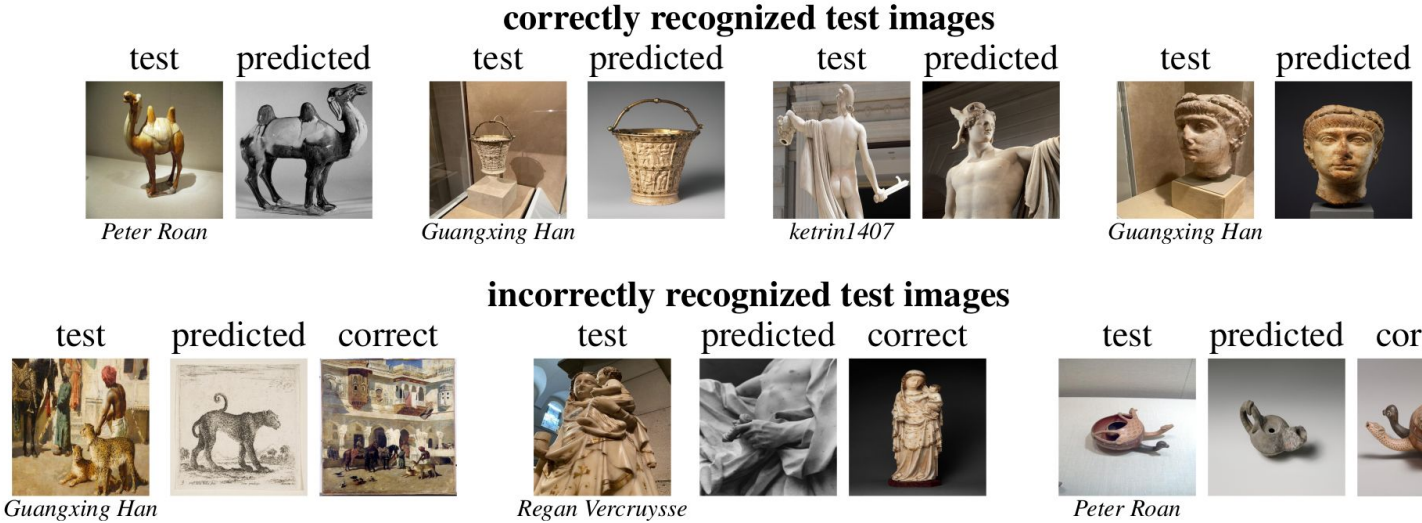
synthetic + real positives

synthetic + mined real positives

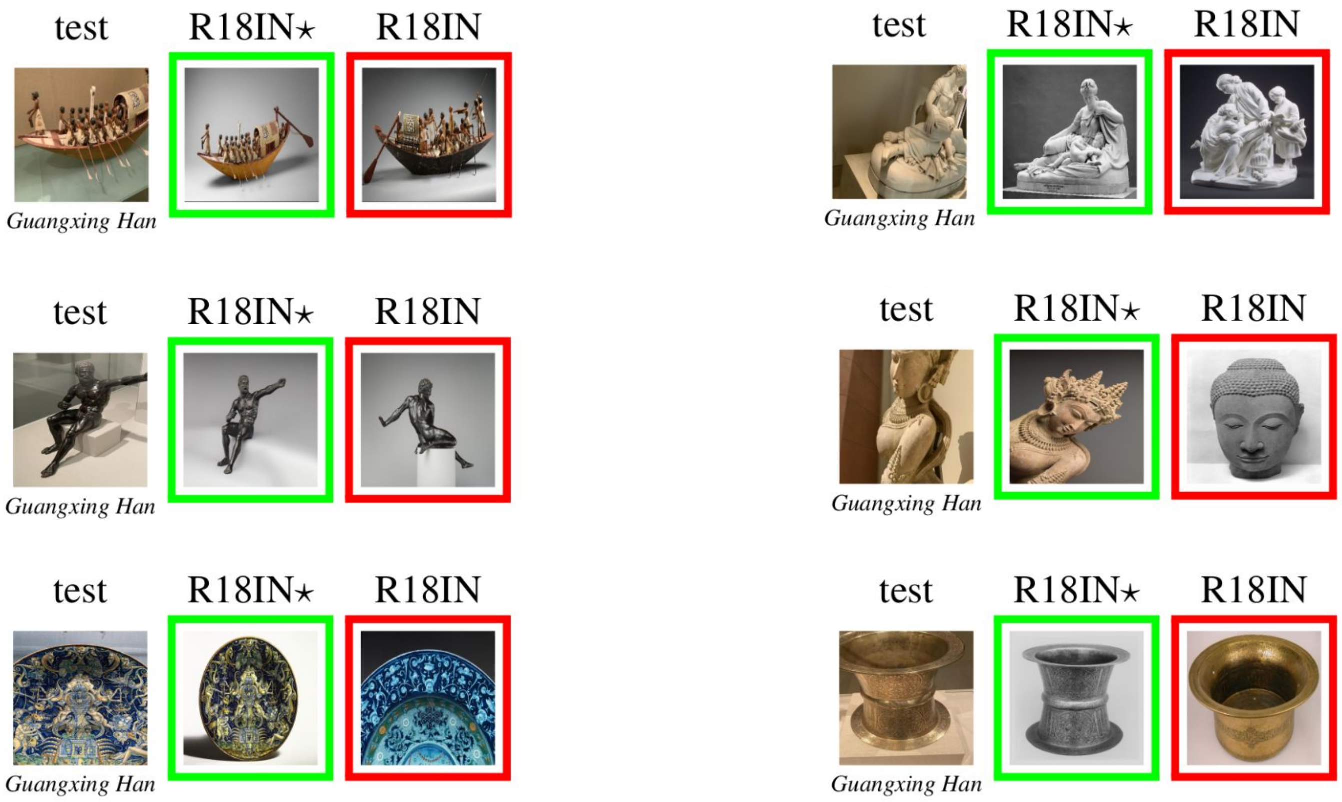
ACC



Qualitative examples



Pre-trained model (R18IN) vs training on MET (R18IN★)



Conclusions

- New dataset on large-scale ILR of artworks
- Complements existing ILR datasets
    - Support future research: generic ILR methods applicable to many domains

- Artwork related pre-training is not necessary useful
    - ILR-related pre-training is more relevant

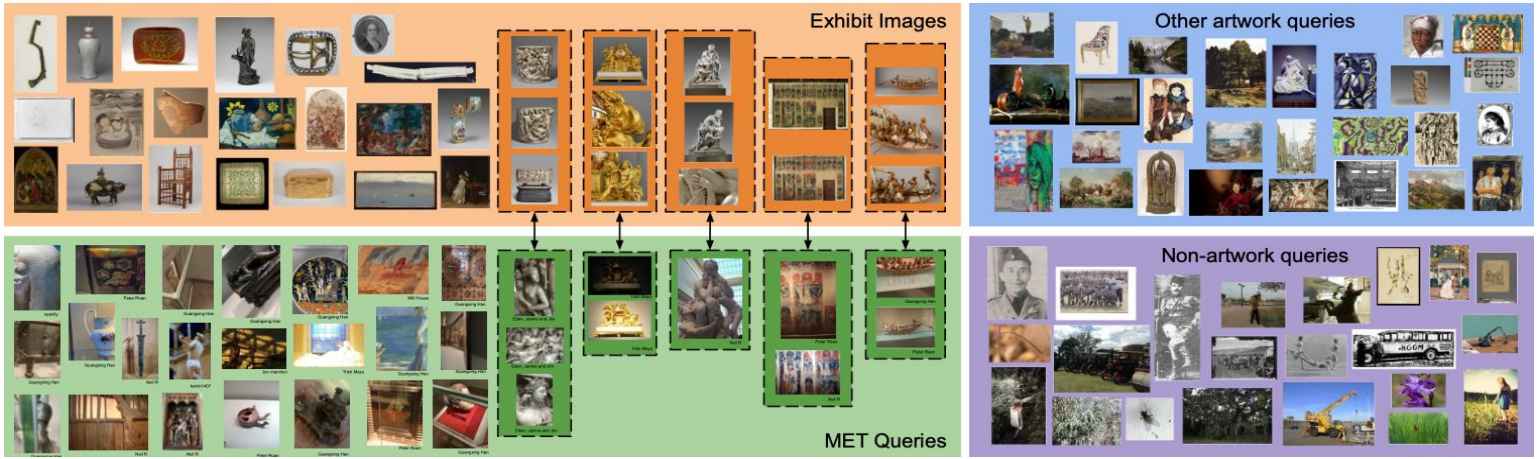
- Superiority of non-parametric vs parametric classifier
- Learning on the MET training set is challenging
    - Relevance of self-supervised representation learning

Dataset page: http://cmp.felk.cvut.cz/met/

Code: https://github.com/nikosips/met

Paper: "**Instance-level Recognition for Artworks: The MET Dataset**"

NeurIPS'21 - Datasets and Benchmarks Track



Thanks to
Andre Araujo, Tobias Weyand, Xu Zhang
MET employees Jennie Choi, Maria Kessler
38 Flickr photographers

Junior Star GACR, grant No. GM 21-28830M